DOCUMENT RESUME

ED 300 418 TM 012 416

AUTHOR Arkley, Harriet; And Others
TITLE Assessing Student Performance for School Improvement.
INSTITUTION Springfield School District 186, Ill.
SPONS AGENCY Illinois State Board of Education, Springfield.
PUB DATE Jan 88
NOTE 108p.
PUB TYPE Guides - Non-Classroom Use (055) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; Criterion Referenced Tests; Educational Assessment; *Educational Improvement; Elementary Secondary Education; *Evaluation Utilization; Item Analysis; Norm Referenced Tests; *School Districts; Scoring; Test Construction; *Testing Programs

ABSTRACT

This handbook exists to assist the staff of a Springfield, Illinois school district in implementing the district assessment program and to assist other local school units in developing and assessing an implementation program. Chapter 1 presents information on the philosophical and pragmatic background of comprehensive assessment programs, with suggestions regarding the planning of such systems and specific details relating to the use of norm-referenced achievement tests. The second chapter presents information on three basic types of locally designed assessment instruments: criterion-referenced tests, performance scales and check lists, and standardized direct writing samples, with a sample of each provided in the appendices to this chapter. In addition, the chapter describes and recommends a particular test development process. A third chapter discusses the development of item analysis and scoring assessment instruments, with suggestions for hardware and software. The final chapter discusses the use of assessment information obtained for school improvement. It is concluded that the assessment of student learning is most effective when it is integrated with curriculum and instruction, and its greatest impact on school improvement occurs when the assessment is managed by building-level staff. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED300418

ASSESSING STUDENT PERFORMANCE FOR SCHOOL IMPROVEMENT

SPRINGFIELD PUBLIC SCHOOL DISTRICT 186

Dr. Donald Miedema, Superintendent of Schools

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D.G. Gill

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Development of this manual has been funded as a Model Practices Project by the Illinois State Board of Education.

TM 012 416

The following staff members of the Department of Instruction, Springfield (Illinois) Public School District 186 are responsible for the production of these materials and the workshops which have been presented as a means of dissemination.

Dr. Harriet Arkley
Ms. Grace Curry
Mr. Donald Davis
Mr. Gregory Dietel
Dr. David Estrop
Dr. Robert Hill
Mr. Michael Holinga
Ms. Kathryn Ransom

Questions and inquiries should be addressed to:

Dr. Robert Hill, Director
Department of Instruction
Springfield Public Schools
1900 West Monroe Street
Springfield, IL 62704
217/525-3026

First Printing, January, 1988

BEST COPY AVAILABLE

Acknowledgements

Sincere appreciation is extended to Mr. Richard Karman for his work in the development of the computer software which has made possible the management of significant amounts of assessment data by the staff of Springfield Public Schools. Without the high quality of his work, this project would not have been possible.

Those who produced this handbook also wish to thank Dr. Thomas Kerins and the staff of the Student Assessment Section of the Office of the Illinois State Board of Education. Their expertise and technical assistance has assisted not only in this project but in the overall efforts of school personnel throughout Illinois to improve assessment efforts.

Gratitude also is extended to the Illinois State Board of Education for funding made available through the Model Practices Project. This handbook and the workshops used to disseminate related assessment information have been funded by a grant award of the project.

TABLE OF CONTENTS

Chapter 1 **COMPREHENSIVE ASSESSMENT SYSTEMS**

Introduction I-1
Philosophical and Pragmatic Backdrops I-1
Vocabulary of School Assessment I-4
Planning a Comprehensive Assessment System I-6
- Purposes of Assessment I-7
- Curriculum Alignment I-9
- Staff Input and Ownership I-10
- Assessment Components I-11
- Timeframes and Volume of Assessment I-15
- Information Management I-16

Norm Referenced Assessment Systems I-17
- Criteria For Judging Achievement Tests I-17
- Advantages and Disadvantages of Standardized Tests I-19

Summary I-20
References I-21

Chapter 2 **LOCALLY DESIGNED ASSESSMENT INSTRUMENTS**

Introduction II-1
Types of Locally Designed Instruments II-2
- Locally Designed Criterion Referenced Tests ... II-2
- Standardized Direct Writing Samples II-4
- Performance Scales and Check Lists II-4

Test Development Process II-6
- Establish a Timeline/Appoint Committee II-6
- Review or Design Objectives by Staff II-6
- Review Research to Ascertain Compatibility with Objectives II-8
- Large Group Review II-9
- Develop Test Items and Standards/Criteria II-9
- Conduct Bias Review of all Materials, and Amend II-11
- Pilot Test II-11
- Test Administration Policy for Consistent Results II-13
- Conduct Second Bias Review II-14

Establish a Standard Criteria for Passing II-14
Develop Reporting, Remediation Procedures, and Staff Development II-14
Administer, Score, and Analyze Tests II-15
Report to Public and Staff II-15
Summary II-15
References II-16
Appendices II-17-II-26

Chapter 3 **ITEM ANALYSIS AND SCORING ASSESSMENT INSTRUMENTS**

Introduction III-1
How to Develop Your System III-2
Development of a Testing Strategy III-3
Development of Criterion Referenced Tests III-4
How to Run the Springfield CRT System III-6
System Operation III-6
Backup of Data III-13
How to Use the Results III-14
Pretest III-14
Posttest III-17
Combination of Pretest and Posttest III-19
How to Cope With Problems III-21
Tentative Timelines III-23
Roles of Various People Associated With the System III-24
Administrators III-24
Teachers III-25
Clerical III-26
Board of Education III-26
Materials You Will Need III-28
Hardware III-28
Software III-28
Appendices III-29-III-33

Chapter 4 **USING ASSESSMENT INFORMATION FOR SCHOOL IMPROVEMENT**

Introduction IV-1
Program Evaluation for School Improvement IV-1

District-level IV-2
Building-level IV-3
Classroom-level IV-4
Planning For School Improvement By Profiling IV-5
What is profiling? IV-5
Who does the profile? IV-6
What is the process? IV-6
A final word about profiling! IV-13
Summary IV-13
References IV-15

Glossary V-1

COMPREHENSIVE ASSESSMENT SYSTEMS

Introduction

This handbook has been created with two purposes in mind. The first is to assist the staff of the school district in which it has been developed in implementing that district's assessment program. The second is to assist other local school units in their development and implementation of assessment programs. The suggestions included in the following pages reflect both the available research base relating to assessment and the practical experiences of school district personnel gained during the design and implementation of a comprehensive assessment strategy.

While issues have been raised and probable answers to questions about assessment have been offered in this handbook, those who have prepared it believe that raising the issues was as important, or even more important, than offering "pat" answers. Schools and school districts differ widely in terms of the needs of students, the demands and expectations of the communities they serve, and the range of issues to be considered when evaluating program effectiveness. For this reason, we believe that locally designed assessment programs will best meet the needs of the students, teachers, and publics of those schools and school districts when the assessment program is assembled by those who will use it and ultimately make decisions based on the results of those assessments.

Philosophical and Pragmatic Backdrops

An obvious assumption related to building an assessment system is that it will integrate with a parallel set of identified learner outcomes (goals, performance objectives, etc.). There exists, however, an even more basic place to begin in designing curriculum and assessment. School assessment is likely to have its greatest positive impact when it is utilized as one component consistent with an overall **belief system** that guides the operation of a school or

school district. Putting a local assessment system in place only in order to comply with state mandates **is not** likely to produce school improvement unless the school or district builds the program on a foundation in which its staff, board members, and parents believe and value as important.

The impetus for assessment programs has grown out of two related phenomena. First, the demand for the **accountability of schools** creates the implication that measuring and reporting student performance is requisite to that accountability. Without question, student test scores have become one of the yardsticks against which school accountability is measured. Putting aside the many legitimate issues surrounding testing (e.g. how much, how often, when, what kind, item bias, etc.), the pragmatic political fact for local schools in the 1980's and 1990's is that assessing student progress and reporting results has become a "given" closely related to the institutional well-being of public schools.

The second stimulus to the movement for school assessment has grown out of the **school effectiveness research** produced during the past two decades. This body of research provides the "high ground" upon which schools and school districts may approach the assessment of student performance as a piece of the foundation of meaningful school improvement. This body of research now has a long enough history that a second generation of it is beginning to emerge. Even though this second generation of research includes some criticisms of the original formula of school effectiveness, this new body of information does still include elements (e.g. academically rich programs, shared governance, personal attention to students, and teaching aimed at preventing academic problems) which require the support of high quality student assessment (Stedman, 1987).

As noted previously, fulfilling the goals of this approach to school improvement requires the development and use of a school **belief system**. Such a belief system must truly reflect those points of view which influence the actions of staff, students, parents, and members of boards of education. The belief system must be founded upon principles that operate every day of the school year throughout all components of the school institution, rather than ones that are

occasionally surfaced at opportune times and to which lip service is paid for the purposes of political expediency.

Considering the elements of the belief system of a school is an important first step to making the assessment program fit with the other operational components. Although there are many issues to consider in coming to grips with the belief system of a school, we recommend that those attempting such a task consider at least the following:

1. The research of Ron Edmonds suggests that effective schools are not staffed by people more virtuous than those in less effective schools.
2. The research of Lawrence Lezotte suggests that excellence is a *state of becoming rather than a state of being.*
3. School improvement, as indicated by measures of student performance, has occurred in schools where all staff believe (and practice this belief) that *all* students can learn. Such a belief is practiced when high expectations are held for all students *equally*. This practice does not imply having equal expectations for every student, but rather it means that all staff hold high expectations for all students regardless of their backgrounds, skill levels, or records of previous school performance.
4. Effective schools utilize information about the prerequisite skills of students to structure instruction in a way that maximizes learning for as many students as possible. In doing so, these schools avoid the practice of "sorting and selecting;" acts which greatly endanger instructional equity.
5. Effective schools do not place much hope in "quick fixes."
6. Decision-making is shared in effective schools. School improvement requires the unleashing of "expert and information" power. School improvement efforts center in and are focused on the school unit.

As local schools and school districts approach the task of developing and using a comprehensive assessment system, it is essential that all involved recognize their roles and the roles of the other participants in the process. A beginning point for all involved is

to recognize that the assessment system will best produce results if its central focus is on school and classroom units. This point carries the implication that **teachers and principals will ultimately have the greatest impact on change, whether it be positive or negative**. Boards of education, superintendents, and other central office staff may be in the position to either impose assessment systems **or** to enlist the support and participation of building-level staff in developing assessment programs. The latter option is much more likely to produce commitments to better schools. The "bottom line" is that board members, superintendents, and central office staff are in a position to support a climate for school improvement and to allocate the necessary resources for effective school programs, but ultimately **the controllers of success or failure will be teachers and principals**. In recognizing this point, it is important that steps are taken to provide both the **power** and **authority** to those charged with the basic levels of responsibility for effective schools so that they may accomplish the school's mission.

A more comprehensive discussion of the roles of various participants in the processes of developing and implementing assessment plans is included in the *Illinois Assessment Handbook* (1987) published by the Illinois State Board of Education. Specifically, the table in Chapter 2 of that document may be of benefit to groups dealing with the issues surrounding the roles of various participant groups.

Vocabulary of School Assessment

Assessment, the topic of this handbook, can be a difficult concept with which to work. One reason for this problem is that the terms assessment, test, and evaluation are often used interchangeably. Because all of these terms are so closely related and because other parts of assessment vocabulary may not be familiar to every reader, we have taken the time to clarify the meanings of five terms that comprise the foundation of this handbook. In addition, a more comprehensive list of terms is offered as a glossary found at the end of the handbook. Understanding the relationship of these five terms,

sometimes marked by their similarities and sometimes by their differences, is an important first step to using the handbook.

1. **Assessment:** Of the three terms assessment, test, and evaluation, the first is the most encompassing. Assessment refers to *any of a number of methods that may be utilized to gather information about the performance of students.* Tests, in fact, are one component of assessment, but many additional means are available for supplementing performance data. The additional components of assessment beyond tests may include observations, interviews, rating scales, artifacts (artwork, term papers, etc.), grade analyses, and background investigations. The term **assessment program** generally includes the processes of planning, designing, administering, and interpreting assessment data. It should be noted that the quality of an assessment program will be very dependent upon the quality of each of its component parts. Specifically, it will be impossible to have an effective assessment program if it consists of test instruments which are unreliable, invalid, or of poor quality.
2. **Test:** The term test is much narrower in implication than either assessment or evaluation. A test typically "denotes the presentation of a specific set of questions to be answered, a task to be performed, or a problem to be resolved. A test is tangible and structured, and it can be administered within a relatively limited period of time." (Alaska Department of Education, 1986). Generally, tests are categorized either as norm referenced or criterion referenced. Each of these two terms is defined in detail below.
3. **Evaluation:** This term refers to *what one does with the information gathered in an assessment program.* Once assessment data are collected, it is assumed that educators will interpret their findings and ultimately base decisions on value judgments supported by the assessment results. It should be noted that program evaluation has many components, one of which is interpretation of assessment data. In this handbook, discussion of data interpretation is covered extensively in the fourth chapter.
4. **Norm Referenced Tests:** Standardized tests, available commercially and administered in nearly every American school district, for the most part are norm referenced. By design, NRT's are intended to be used *to compare the performance of a student or a group* of *students to the performance of other students.* The undergirding foundation of NRT's is the statistical concept of normal distribution. The underlying assumption of NRT's is that the scores of all

students taking the test will be distributed in a pattern matching the bell-shaped or normal curve. For this reason, NRT's as they are used in school assessment programs, include questions of a low difficulty level, others of a more moderate difficulty level, and yet others that very few students are expected to answer. The discriminatory nature of this range of questions is important to distributing performance scores. Examples of NRT's include commercial achievement tests, the ACT, the PSAT, the SAT, and the tests currently being developed by the Illinois State Board of Education. Test scores on NRT's are typically reported as percentile scores, scale scores, and grade equivalents.

5. **Criterion Referenced Tests:** Criterion referenced tests differ from norm referenced tests in that the performance of the *test taker is judged against a fixed standard based upon a predetermined set of learner objectives* rather than against a statistical standard based on the normal distribution of test scores. On a CRT, each test item is developed to determine some level of mastery on the part of the test taker on one or more learner objectives to which the test item is matched. The scores on a CRT are related in some way to what a student specifically knows or can do. In an objectives driven instructional program, the desired outcome will be to have all students master all items on the CRT once they have been engaged in the appropriate learning activities.

   Many publishers now make available CRT's to accompany their instructional materials or other curricula. Local school districts, as well, have developed CRT's of their own tailored to the learner objectives and expected outcomes of their instructional programs. It should be noted that most major publishers of norm referenced **achievement tests** also make results available to their customers which outline criterion referenced results on an item-by-item basis. (These results provide information about how well students mastered the objectives, first as a comparative measure in reference to other students and secondly as a measure of how well specific objectives were mastered.) In this regard, one commercially prepared test may be utilized in two different ways. The one precaution necessary, however, when one considers using achievement tests as CRT's is that the number of test questions per objective is limited. This factor must be taken into account when interpreting test results.

Planning a Comprehensive Assessment System

There are six factors to be considered in mapping out a comprehensive assessment plan. It is important that local school

districts include in their planning a means to address each of the six factors in a coordinated way. The elements of the planning process overlap, both philosophically and pragmatically, so it will be necessary for the personnel managing the planning to keep all of the elements in perspective throughout the process. The six factors which comprise the planning for a comprehensive assessment plan include:

- Purposes of Assessment
- Curriculum Alignment
- Staff Input and Ownership
- Assessment Components
- Timeframes and Volume
- Information Management

Purposes of Assessment

As previously noted, the term assessment refers to more than tests and testing. Consequently, a comprehensive assessment system will include a number of components, some of which will not be tests. In piecing together the components of such a system, school personnel will want to keep in mind three basic purposes for all assessment. In doing so, the designers of the assessment system will want to be able to determine how each component of the system addresses one or a combination of the three basic purposes. An understanding of the purposes of assessment will aid school staffs in reaching decisions about which students to test, what content to assess, what instruments to use, and when to use them.

The first purpose of assessing student performance is to use the assessment as **one part of an overall instructional plan**. Many models of lesson design exist, but each contains an element which signifies the importance of assessing student learning. The assessment step in the lesson design reinforces the need for review, and it serves the purpose of causing students to draw upon previously learned skills and information in applying new ones. The majority of these types of assessment will be teacher designed and/or teacher directed, but the possibility does exist to fit this type of assessment with either one of the other two purposes of assessment, both within a single classroom or on a school-wide or district-wide basis. When students know and recognize assessment as part of the overall instructional approach, a specific value of importance is

communicated about assessment from the teacher to the student. In a similar manner, an organized assessment system operating in a school building or school district can also communicate a certain statement of value to teachers, parents, students, and the community that can enrich the learning climate.

A second purpose of assessment is that of **diagnosis and prescription.** Assessment measures can be used effectively to determine student's prerequisite skill levels and previously learned information. Assessment used in this manner will aid teachers and other school staff in determining the strengths and weaknesses of individual students or groups of students. Such information can then be utilized to assist curricular planning and pacing. In an objectives driven instructional program, for example, criterion referenced tests administered at the beginning of a course or unit (pretest) can be used prescriptively to provide information to the teacher that s/he can utilize in many ways. This type of information is valuable in determining the grouping of students based upon their prerequisite skills, or it can help a teacher determine which concepts of the course will be review for students and which will represent new information and skills.

The point has already been made that the various purposes of assessment can be combined in some instances. For example, a teacher could administer a test or use a check list of skills as part of his/her lesson design strategy after having taught the material and after having provided guided and independent practice. In this way the teacher would require the students to engage in review of the material. Upon scoring the assessment instrument, the teacher could then use the results to determine if reteaching were necessary, if the regrouping of students were necessary or desirable, or if going on to the next unit or set of skills were possible. In this way the same assessment becomes diagnostic and possibly prescriptive. Assessment instruments utilized in this way can be teacher-made, or they can be provided by the school or district as part of a comprehensive assessment program.

The final purpose of assessment is the one most easily recognized and frequently used. An obvious reason for assessing student

performance in school is to conduct **evaluation**. At the level of interaction of teacher-to-student this function most often represents assigning a grade to the work performed by a student. This type of activity occurs frequently and periodically in most schools.

Assessment for evaluation, however, does have a broader context than teachers assigning grades to students work. Any element of a building-level or district-wide assessment program that produces information upon which a judgment of merit can be based is evaluative in nature. This point is at the foundation of the school accountability movement. The school effectiveness research seems to indicate that effective schools consistently practice decision making by considering an assessment of student performance measured with a "value base" (the reading program works, students can add to our satisfaction or they cannot, the graduation rate is acceptable or unacceptable, etc.).

The project team responsible for this document wishes to make an important point about assessment for evaluation. That is, far too many educators see this function of assessment as the only one. It is quite easy to utilize the same assessment instrument to promote review, to check the level of learning and proficiency of students (diagnosis), and to give grades. What frequently happens, however, is that educators give the test, assign the grade, and go on to the next unit **even if some or many students do not exhibit competence on the unit just tested.** We believe that an effective assessment system will help foster attitudes and practices which result in all staff analyzing assessment results from the broader perspective outlined here. An even more comprehensive system will offer technical support to teachers and principals to use in conducting such analyses.

Curriculum Alignment

As noted previously, a match between instructional objectives and the assessment of them is basic to creating an effective assessment system. The concept of curriculum alignment implies a process of matching three components that are intended to produce student learning. They are **curriculum, instruction, and**

assessment. The effective and efficient operation of each of these components is not likely to occur in isolation. Consequently, school personnel will have to consider the interrelationships of the three during efforts to build and maintain comprehensive assessment systems.

Curriculum alignment is a topic deserving of an entire handbook of its own. (See Chapter 4 of the *Illinois Assessment Handbook* published by the Illinois State Board of Education for a more detailed discussion of the topic.) It is not possible in this document to detail all aspects of curriculum alignment, but two very general points about its relationship to assessment are offered.

1. An aligned system which successfully integrates curriculum, instruction, and assessment will depend on **accurate and frequent communication** among teachers and administrators at all levels of the system. The use of communications links to purposefully share information about alignment will diminish the opportunities for misunderstanding, blaming, and "buck-passing".
2. The need to accurately define the alignment of curriculum, instruction, and assessment implies the need to put the plans **in writing**. Written goals and objectives, instructional strategies, and assessment plans **do not** guarantee a well articulated and aligned system, but their absence does suggest that little chance exists for all of the educators in a system to understand their own roles and responsibilities or those of their colleagues. In such a case, adequate alignment is not likely to occur (Alaska Department of Education, 1986).

Staff Input and Ownership

The involvement of all segments of the professional staff of a school or school district will be as vital to effective assessment as it is to curriculum and instruction. Current school effectiveness research is producing detailed information about a topic frequently defined as "shared management" or "shared decision-making." The findings of this research suggest that school improvement is most

likely to occur when policy, practice, and procedure are shaped by the contributions of a broad base of the professional staff and when the responsibility and authority for producing effective schools are placed collectively in the same hands.

Imposed assessment programs, handed down from one hierarchical level to another, may result in compliance on the part of school staffs, but such programs **are not** likely to produce the desired results of school improvement. One reason that success is unlikely in such a situation is that those at the lower levels of the hierarchy (probably teachers and principals) will feel threatened rather than involved. In addition, it is likely that these staff members will not receive support from an imposed system that best meets **their** needs at **their** levels of operation.

In recognition of the possible negative factors described above, those responsible for this document strongly urge that comprehensive assessment systems be built with the input of teachers and principals. Furthermore, we support the idea that use of the information produced in such a system should and must be useful to these staff members. In the fourth chapter of this handbook the use of "school profiles" by "school improvement teams" is discussed. In other portions of the handbook continual reference is made to the need for the input of teachers and principals in the development of objectives and assessment instruments.

Assessment Components

The components of an assessment system should be closely related to the purposes of assessment as defined by a school unit. In developing a comprehensive assessment system, school personnel will need to identify the various needs and purposes to be addressed with the use of the system before reaching conclusions about its components. The decisions about which components to include in a local system most likely will be influenced by instructional priorities, the ability to manage the information produced in the program, available resources (both fiscal and human), state mandates, and local expectations. The blending of these various considerations into a decision making process will in many instances

require the establishment of priorities, particularly if identified needs exceed available resources.

No formula for the perfect assessment program exists. Based upon their experiences and review of the local practices of many school districts the project staff responsible for this document has developed the following list of general types of components school units may want to consider in developing their own assessment programs. We believe a comprehensive program will not depend entirely on any one component, but rather it will include a blending together of several or all of them.

1. **Achievement Tests:** Achievement tests represent a school's window to the world. The norm referenced group data produced in an achievement test allows a school or district to compare the performances of its students to those of students from many locations. The types of information produced in achievement tests provide school units with the opportunity to decide if their curriculum and instruction matches with national practices and trends. Without this type of information, school units might be able to determine the performances of students in respect to local expectations, but those local expectations could differ greatly from standards which their students might have to attain in order to gain admission to college or to the work force.

   The publishers of many achievement tests today also provide test results which can be used as criterion referenced data. Depending upon the degree of match between the items contained in the achievement test and the instructional program of a school unit, achievement test results can provide valuable criterion referenced data to schools.

2. **College Entrance Exams:** These types of norm referenced tests also provide valuable information to local school units. Publishers of the ACT, SAT, and PSAT programs make available to local school units a wide range of information which can be useful in determining the relative success of the school unit in preparing students to enter college and university programs. In addition to the scores on these tests which provide a means for comparison with students in other school units, the publishers also make available performance reports that permit analyses of the areas of strengths and weaknesses of a school's students on these

types of examinations. Such information can be useful to the process of curriculum design and revision.

3. **Commercially Available CRT's:** It is possible for school units today to purchase a variety of commercially prepared criterion referenced tests. CRT's which are available commercially vary in their potential usefulness to local school units depending on their match to curriculum and their relative costs. The most common type of product available in this category is a testing supplement prepared by the publishers of textbooks. Typically, these products offer the appeal of providing a match to the instructional program of a school if the companion text materials also are adopted.

   Other types of commercially prepared CRT's also are available in the market place. These products range from single test instruments to sophisticated sets of learner objectives supported by test item banks and computer operated management systems. These types of materials and systems tend to be expensive.

4. **Locally Developed CRT's:** Many local school units have initiated the practice of designing and implementing their own criterion referenced tests. The impetus for this practice began in many locations with the movement toward competency based instructional programs. Such an approach demands a CRT system in which the progress of all students can be tracked. An advantage to using locally designed CRT's is the built-in match of assessment to the other components of curriculum alignment. A second strength of the approach is that the development and refinement of assessment, curriculum, and instruction become mutually reinforcing. At the same time, this type of test development traditionally includes significant levels of teacher and principal input and ownership. The use of these types of tests often produce the need in a local school unit to have a system for managing large amounts of test data. Extensive discussions of both the development of these types of test and their uses are included in the second and third chapters of this document.

5. **Writing Samples:** Student writing samples represent a specialized type of criterion referenced assessment tool. The type of sample developed by the Illinois State Board of Education requires students to write a brief composition in response to one or more of a variety of writing prompts (short stories or scenarios). Each student's writing is then scored against a previously established

set of criteria. A numeric score is assigned to the student's work within each of the categories defined by these criteria. It will be possible, once this type of writing assessment is used more universally, to norm the scores of students and thus provide for comparisons of groups of students. This type of use of the writing scores is anticipated in Illinois.

6. **Performance Scales and Check Lists:** Some areas of student performance are not easily assessed with the use of traditional test instruments. This fact, however, in no way diminishes the relative importance of these areas of student performance or the need for schools to be able to measure student progress. Examples of school programs in which the use of performance scales and check lists have been developed include music, the visual arts, and physical education. One advantage of the use of performance scales and check lists is their easy integration into the instructional program. Some school districts using performance scales and check lists have begun to develop ways to integrate the assessment data produced with their use into the existing management system of the assessment program.

7. **Artifacts:** Artifacts represent tangible evidence of levels of performance attained by students. Examples of artifacts include works of visual art, term papers, original writings, tape recordings of speeches or dramatic productions, minutes of meetings, agendas, etc. Other components of an assessment system can be used to rate or score artifacts (e.g. performance scales or grades) , or they may be utilized as part of assessment based solely on their developed existence. For example, the records of a school club that confirmed the contribution of 1,000 hours of volunteer service to the community could be used as part of the assessment of the goals of citizenship education.

8. **Grade Analyses:** It would be difficult to find a school system in which some types of grades were not assigned to the work of students. It is likely that the grades assigned by teachers to the work of their students, in and of themselves, include many of the components of the comprehensive assessment program described in this document. And, yet, it probably also is true that few school units directly incorporate the grades assigned by teachers into program evaluation or school improvement efforts. The thoughtful analyses of the grades assigned to the work of students by their teachers, based on long-term and

frequent interactions, should be considered as part of the comprehensive assessment system of a school unit. The great advantage of using this information is that it already exists, and retrieving it probably does not require extraordinary effort or expense. Furthermore, to ignore the use of grade analyses may send an undesirable message to teachers. That message is that program assessment is some externally imposed process rather than one that is integrated with the instructional program. The intent of this handbook is to portray quality program assessment as quite the opposite of a periodic imposition.

Table 3.1 found in the *Illinois Assessment Handbook* (1987) contains a list of advantages and cautions relating to various types of assessment instruments. That table provides additional beneficial information which local school units can utilize as they consider what types of instruments to include in their assessment programs.

Timeframes and Volume of Assessment

The issues of when to assess and how much assessment is enough (or too much or too little) are critical to the planning of an assessment system. In terms of complying with state mandates, the timeframe question is substantially answered for Illinois school districts. That timeframe says that assessment must occur by April of each school year, and that improvement plans based upon the results of that testing must be submitted by the following November. Without question, compliance with the mandates will be a factor in planning time sequences for Illinois school districts.

Each local district, however, has been left some degree of autonomy in the design of its overall assessment plan. Conducting that planning will require thoughtful consideration to the timing of assessment in a way to produce useful and usable information.

The question of when to assess is closely related to the question of how much is enough. If a school unit decides to utilize multiple components in an assessment strategy, the need to spread that assessment across more grade levels and more months of the school

year is increased. Those who prepared this handbook believe that such a system is desirable, so we also believe that the assessment will be most effective if utilized over a broad period of time with the involvement of students at a variety of grade levels. Such a system, we believe, is possible within the guidelines of Illinois school reform.

Many legitimate concerns and questions have been raised about balancing the need for assessment with the amount of available instructional time. Those concerns most often arise if one considers assessment as a subsidiary to instruction rather than as an integral part of it. We have offered as suggestions many different components of a comprehensive assessment plan. Some of those components can be integrated as part of the instructional program in ways that do not require taking away from teaching time. We encourage local school units to consider assessment as a critical component of the instructional program, and in doing so to plan ways to incorporate assessment as part of curriculum alignment rather than as a separate entity which is dropped in on occasion.

Information Management

School personnel have been known to say that collecting assessment information is the easy part of the process. The more difficult part of the equation is managing the information in ways that support accurate evaluation, policy formation, and program planning. If a school unit initiates the first five components of a comprehensive assessment plan (defining purposes, aligning curriculum, getting staff input, using multiple components, and designing timeframes), the only guarantee that can be made in the absence of the sixth component is that the school unit probably will have more information at its disposal than it can use. For this reason, it is vitally important that school personnel devise a means to manage the assessment data it will collect.

The third and fourth chapters of this handbook offer illustrations of information management systems. The models presented are intended to be of practical use to school units of all sizes and differing student compositions.

Norm Referenced Assessment Systems

All of the components of a comprehensive assessment system are discussed in significant detail in the remaining chapters of this handbook, except for norm referenced assessment systems. Because achievement tests are widely used by local schools as part of systematic assessment, discussions of two factors relating to their use are included as the conclusion to this chapter. Specifically, information regarding criteria for judging achievement tests and a listing of advantages and disadvantages are offered.

Criteria For Judging Achievement Tests

When deciding which achievement test to use, school personnel should realize that no perfect test exists. On the other hand, however, if the steps described in this handbook for determining the belief systems of the school unit have been taken, the chance does exist to obtain an achievement test which matches well with the needs and values of that school unit. The Alaska Department of Education in 1986 produced an outstanding guide to assessment issues. Included in that guide are seven criteria for judging achievement tests. Collectively or individually these seven criteria should assist local educators in addressing the technical, practical, and alignment issues related to test selection. Each criterion is discussed below.

1. **Content validity:** The issue of validity relates to whether or not the test measures what it purports to measure. In evaluating an achievement test, local personnel will want to determine whether or not the questions on the test that assess students' reading abilities (or math, or study skills, etc.) match with what local teachers believe represents the ability of students to read. If that match exists, the test has the desired content validity. If that match does not exist, then the test probably would not be appropriate for use in that school setting. As noted before, however, it may not be possible to find a perfect match on any available test. In that case, school personnel will have to seek the best available match.

2. **Co-normed measure of scholastic aptitude:** Most major test publishers offer a companion test of aptitude to accompany their achievement tests. The coordinated use of the aptitude test and the achievement test makes possible the

comparison of the performance of a group of local students with that of a group of students from the national population displaying similar aptitudes on the co-normed test. Some test publishers produce anticipated scores for students based upon their performances on the co-normed aptitude test. In districts making use of these types of information it is possible to evaluate program effectiveness in relationship to the concept of **what value was added** to the learning of students by the school programs in which they have been enrolled.

3. **Norms for individual items:** Using an achievement test that includes norms for each individual test item allows local school units the opportunity to analyze each small component of that test. Doing so may allow the district to identify meaningful differences in the performances of their students (either positively or negatively), or it may assist in the identification of areas of differences with which the district has little curricular concern (areas of mismatch between the test and the curriculum.)

4. **Recently normed:** The age of the norms of an achievement test should be of concern to a local school unit when selecting achievement tests. As norms become older for achievement tests, questions regarding the tests' validity and reliability will increase. Tests on which the norms become outdated most likely will contain sections which no longer reflect current practice or areas of instructional emphasis from a national perspective. For this reason, recently normed achievement tests are recommended.

5. **High reliability:** The reliability of an achievement test is related to whether or not it will produce consistent types of results over the long term. In establishing the reliability of an achievement test, the publisher is attempting to offer schools an instrument that will produce consistent results if the controllable factors of testing are held constant. Because most school units use an achievement battery more than one time, reliability of the test should be a key concern in test selection.

6. **Maximum amount of information in minimum testing time:** This is a "bang-for-the-buck" issue. Instructional time is regarded by most school districts as a scarce commodity. For this reason, school personnel will want to

consider the ratio of useful information produced in an achievement test battery to the amount of time needed to administer it.

7. **Individual report forms:** Before selecting an achievement test, school personnel should have in mind what uses they will make of the test data. With these potential uses in mind, these personnel will be able to determine whether or not the report forms of an achievement test will serve the needs they have in mind for it. Most major test publishers today provide a very wide selection of report forms, so school personnel will have many options available to them when ordering test reports. Because the number and types of reports ordered are related directly to the cost of the test program, identifying local uses of the reports in advance of ordering will greatly aid school units in maximizing what information they get for the dollars they spend.

Advantages and Disadvantages of Standardized Tests

There appear to be relatively few public school units that do not utilize standardized achievement tests in some way. However, the uses that these schools make of achievement tests vary widely. For this reason, a listing of the advantages and disadvantages relating to the use of standardized tests as developed by the Alaska Department of Education (1986) are included below.

This information has been included because we believe every school that uses these types of tests will benefit from the periodic consideration of the advantages and disadvantages. The thoughtful consideration of these issues by local school units is an important step to having the overall assessment program produce information upon which good decisions can be based.

Advantages of standardized tests include:

- These tests contain high technical quality due to the investment of significant resources in their development.
- The content of these tests is representative of what is being taught in classrooms around the country.
- The norms of these tests allow comparisons with external groups.

- Most publishers of these tests provide support and consulting services to local school units who purchase their products.
- Local school units do not have to spend resources to develop these types of tests.

Disadvantages of standardized tests include:

- It is often difficult to get a good match between test content and what is emphasized in the local school unit.
- These tests usually contain too few items per objective to allow the test to be used for diagnostic purposes or for making certain kinds of accountability and evaluation decisions.
- These tests can be very expensive, especially if the school unit desires to have specific kinds of reporting instruments produced by the test publisher.
- The costs of standardized test programs are recurring on an annual basis.

Summary

In the first chapter of this handbook information has been presented regarding the philosophical and pragmatic backdrops relating to the development and operation of comprehensive assessment programs. In addition, suggestions regarding the planning of such systems has been offered, along with specific details relating to the use of norm referenced achievement tests. In the succeeding two chapters of the document, information regarding the local development, use, and scoring of criterion referenced tests is presented. The final chapter of the handbook focuses on the interpretation of various types of assessment data by school personnel with the goal of facilitating school improvement efforts.

References

Alaska Department of Education (1986). *Assessment Handbook: A Practical Guide for Assessing Alaska's Students.* Juneau, Alaska.

Edmonds, R.R. "Programs of School Improvement: An Overview." *Educational Leadership.* 40:3, 4-11, October 1979.

Illinois State Board of Education. (1987). *Assessing Reading In Illinois,* Springfield, Illinois.

Illinois State Board of Education (1987). *Illinois Assessment Handbook* (draft copy), Springfield, Illinois.

Lezotte, Larry (1985). "Effective Schools." Paper presented at Conference on Effective Schools. New Orleans, Louisiana.

Mann, Dale. (1986). Lecture delivered at Administrator Inservice. Springfield, Illinois.

Rowe, Larry. (1987) Lecture delivered at Danforth Foundation Seminar. New Orleans, Louisiana.

Schalock, Del; Fielding, Glen; Schalock, Mark; Erickson, John; and Brott, Michael. "Integrating Teaching and Testing with Program Management." *Educational Leadership.* 43:2, 55-58, October, 1985.

Stedman, Lawrence. "It's Time We Change the Effective Schools Formula." *Phi Delta Kappan.* 69:3, 215-223, November 1987.

Stiggins, Richard J. "Improving Assessment Where It Means Most: In the Classroom." *Educational Leadership.* 43:2, 69-74, October, 1985.

LOCALLY DESIGNED ASSESSMENT INSTRUMENTS

Introduction

A comprehensive assessment system was described in Chapter 1. Locally designed assessment instruments are a part of a comprehensive assessment system. *The purpose of this chapter is to describe the process by which a school or school district might create its own assessment instruments.* In particular, this chapter will focus on the creation of criterion referenced tests, writing samples, performance scales and check lists.

As noted in Chapter 1, the writers of this document believe that a comprehensive assessment system should include locally designed assessment instruments. There are at least two advantages to including locally designed instruments in an assessment system. First of all, locally designed instruments are able to measure the performance of students against locally designed standards that in turn reflect a particular set of objectives and a particular curriculum. In other words, the use of locally designed assessment instruments encourages the alignment of local objectives and assessment. Secondly, this type of test development normally includes teacher input and thus the chances of teacher ownership are increased. It follows that the chances of teachers **using** the test results to revise their instruction or curriculum are enhanced.

Locally designed criterion referenced tests, standardized direct writing samples, performance scales and check lists are defined briefly in Chapter 1, pages 12-15. Comments on the construction of and appropriate use of each of these assessment tools follows.

A sample of a criterion referenced test, standardized direct writing sample, performance scale and check list may be found in Appendices A-E. Certainly one of the first jobs of a school or school district in the process of creating or refining a comprehensive assessment system is to review these and other assessment tools and to determine which of these instruments would be most useful.

Types of Locally Designed Instruments

There are three basic types of locally designed assessment instruments which will be described in this chapter. They include: criterion referenced tests, standardized direct writing samples, and performance scales and check lists. Each is discussed in detail in the sections which follow.

1. Locally Designed Criterion Referenced Tests

The criterion referenced test assesses the performance of each test-taker in relation to a given standard or competency. Success is defined as being able to perform a specific competency or skill, and as a result there is no limit to the number of test-takers who can succeed on a criterion referenced test.

Before constructing a criterion referenced test, it is necessary to know the purpose of the test. There are at least five possibilities (Shrock and Coscarelli, 1986)

1. **Prerequisite tests** -- used to ensure that learners have the background knowledge for the course.
2. **Entry tests** -- used to identify skills taught in a course that the entering student already possesses. The purpose is to allow students to bypass instruction in (or have their instruction limited to brief review) in units whose objectives they have already mastered.
3. **Equivalency tests** -- used to determine whether a learner has already mastered the course's terminal objectives before instruction begins. Typically, the equivalency test is used to decide whether a student can skip a course.
4. **Diagnostic tests** -- used to assess mastery of a group of related objectives in an instructional unit. Whereas the entry and prerequisite test are used before instruction, the diagnostic test will typically be used in conjunction with the instruction or as part of a posttest to determine whether the student is having difficulty.
5. **Posttests** -- given after instruction to assess performance on final objectives.

The same question can often appear on different types of tests. For example, items on the posttest can be the same as or similar to the items on the equivalency test. In fact, these test types are relative -- a prerequisite test for one course could be a posttest for a previous course.

The actual construction of criterion referenced test items or questions and establishing validity and reliability are topics beyond the parameters of this manual. Certainly most of the principles of testing and measurement apply to the design of items on criterion referenced tests. The reader is referred to the references at the end of this chapter for helpful sources.

The team that prepared this manual most frequently uses three types of test items. All three may be scored using a machine scoring system if only one correct response per question is expected. This machine scoring system is described in detail in Chapter 3.

The three types of items frequently used on criterion referenced tests in the Springfield School District are as follows:

1. **Binary-choice**. In a binary-choice item, the student must choose the correct answer from only two alternatives (true-false, yes-no).

2. **Matching**. Matching items require the student to relate or match one list of items with terms or concepts on another list.

3. **Multiple-choice**. A multiple-choice item consists of a "stem" question and a series of possible responses (no more than five if using a standard machine scorable answer sheet) See Appendices A and B for examples.

Deciding on the **number** of items to be used in a test is no simple task. Statistical arguments exist for using between five and 25 test items per objective. As one can imagine, following this advice might result in having a test that lasts longer than the course, so many compromises need to be made.

2. Standardized Direct Writing Samples

As described in Chapter 1, the writing sample requires a student to write a short composition in response to a writing prompt. Trained raters assess the results using a common scale. Standardized prompts and uniform rating scales exist and can be helpful if a school district decides to include writing samples as a part of its comprehensive assessment system (See *Write on Illinois*, ISBE, 1987). For an example of a persuasive prompt, see Appendix C.

The value of this procedure is widely recognized but it is demanding and time consuming. Technical assistance is available through the Educational Service Centers.

3. Performance Scales and Check Lists

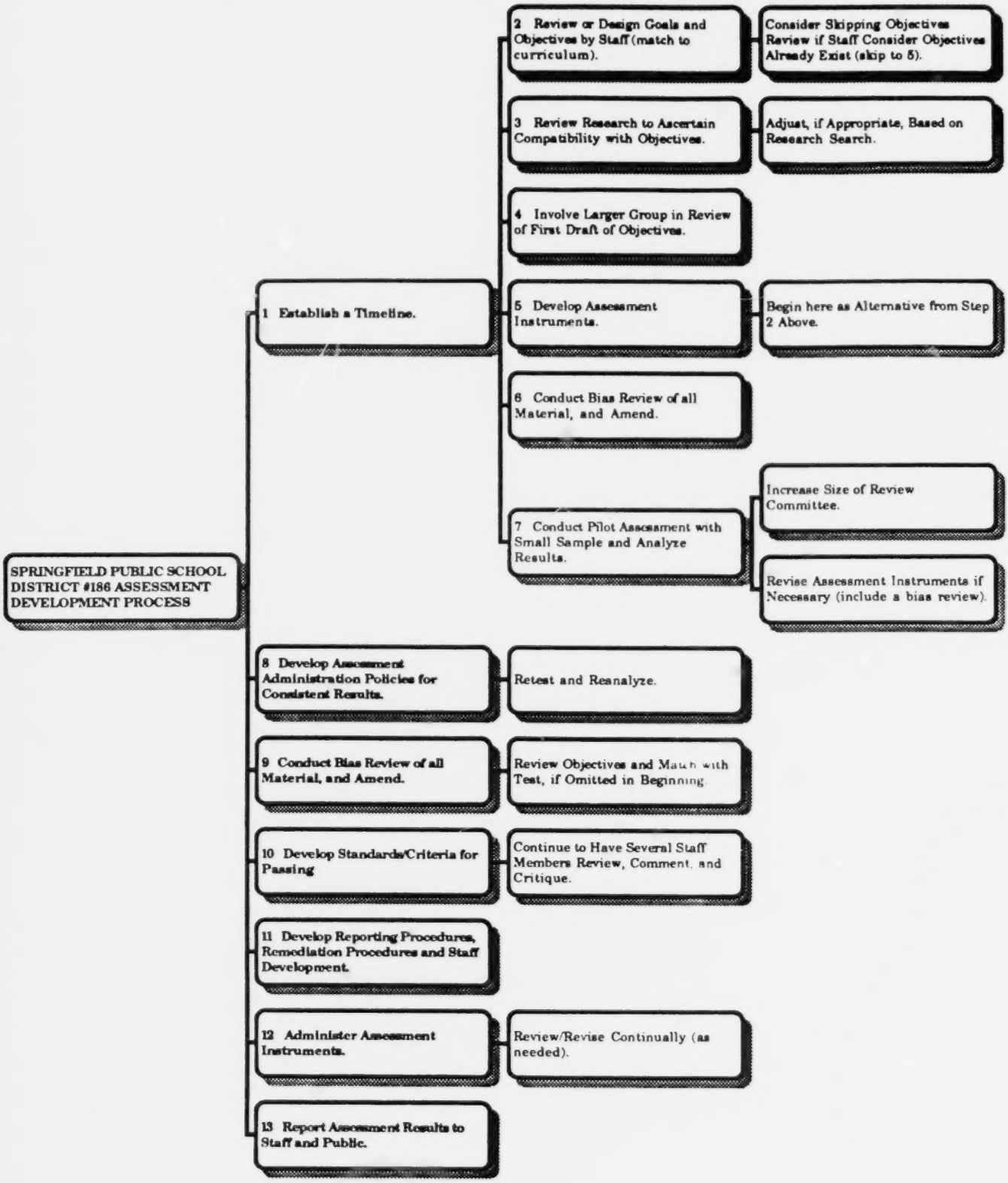
Using performance scales and check lists requires students to demonstrate their achievement in a particular area by performing a specified task(s). A major advantage of this type of procedure is that it may be the most effective method to assess some objectives, especially in the fine arts and in physical education.

Performance scales and check lists need, of course, to be based on clearly stated instructional objectives. The field testing of scales and check lists is especially important as raters need to quickly and clearly understand the items and scales. See Appendices D and E for two examples of check lists.

Three types of locally designed instruments have been described -- the criterion referenced test, the writing sample, and performance scales and check lists. Regardless of which type of locally designed test is used, the writers of this manual recommend using a predetermined test development process. This process is graphically displayed as Figure 1. The remainder of this chapter describes this process.

Figure 1.

SUGGESTED DEVELOPMENT PROCESS
LOCALLY DESIGNED ASSESSMENT INSTRUMENTS


SPRINGFIELD PUBLIC SCHOOL DISTRICT #186 ASSESSMENT DEVELOPMENT PROCESS
1 Establish a Timeline.
2 Review or Design Goals and Objectives by Staff (match to curriculum).
Consider Skipping Objectives Review if Staff Consider Objectives Already Exist (skip to 5).
3 Review Research to Ascertain Compatibility with Objectives.
Adjust, if Appropriate, Based on Research Search.
4 Involve Larger Group in Review of First Draft of Objectives.
5 Develop Assessment Instruments.
Begin here as Alternative from Step 2 Above.
6 Conduct Bias Review of all Material, and Amend.
7 Conduct Pilot Assessment with Small Sample and Analyze Results.
Increase Size of Review Committee.
Revise Assessment Instruments if Necessary (include a bias review).
8 Develop Assessment Administration Policies for Consistent Results.
Retest and Reanalyze.
9 Conduct Bias Review of all Material, and Amend.
Review Objectives and Match with Test, if Omitted in Beginning
10 Develop Standards/Criteria for Passing
Continue to Have Several Staff Members Review, Comment, and Critique.
11 Develop Reporting Procedures, Remediation Procedures and Staff Development.
12 Administer Assessment Instruments.
Review/Revise Continually (as needed).
13 Report Assessment Results to Staff and Public.

Test Development Process

In this section a thirteen-step procedure for test development is outlined. Some flexibility to sequence is possible, but, generally speaking, the process is intended to follow the order in which the steps are presented.

1. Establish a Timeline/Appoint Committee.

In establishing a timeline for designing testing instruments it can be helpful to first determine the completion deadline. This deadline may in fact be set by the beginning of a school year, the implementation of a new curriculum and/or a deadline imposed by the Illinois State Board of Education. After determining this completion deadline one can work backwards, setting dates for the completion of each part of the test development process. There is value in providing ample time for completion of this process for the fermentation of ideas, piloting of tests and subsequent revision of tests is a time-consuming task.

The committee that is assigned to develop test instruments should be balanced in terms of teaching experience, teaching assignment and expertise in test development. The committee needs to be a workable size, probably between five and ten people. The chairperson should have a clear definition of the assignment and the test development process. The chairperson will also need to decide whether the committee will be released to work on school time or be paid to work after school or during the summer months.

2. Review or Design Objectives by Staff (match to curriculum).

Define Goals/Objectives: The initial task of most committees is to determine the goals and objectives for which the assessment is to be designed. This document is not intended to be a comprehensive handbook on writing objectives. There are many other sources available for that purpose.

A few reminders, however, might be helpful. First of all, the more clearly the objectives are defined and the more closely they relate to the actual curriculum being taught, the easier it is to develop an assessment plan. Secondly, staff involvement in the

development of the objectives is essential. There is no substitute for having the actual review of the curriculum. As staff begins this process, they should refer to the *State Goals for Learning and Sample Learning Objectives* in a specific subject area and the sample set of district-level learning objectives which are consistent with those Goals. These Goals are listed in a publication from the Illinois State Board of Education. Although school districts have the option of adopting or adapting these objectives, the local plan must equal or exceed the sample objectives.

Broadly Define Objectives: The learning objectives most appropriate for assessment at a district level will generally be broader than at the classroom level. These objectives will generally represent terminal goals as opposed to intermediate, daily tasks. Thus, one may wish to consider a set of rather broad objectives to meet ISBE requirements and specific behavioral objectives for individual courses offered. For example, these goals might be stated, "At the end of the unit, the learner will perform, demonstrate, write . . . .".

Assistance: Many forms of assistance for developing objectives are available. A set of sample learning objectives is in the ISBE document, *State Goals for Learning*. Ten local school districts were funded in FY 86 as demonstration sites for development of model outcome/assessment systems and a list of those schools could be obtained from the Illinois State Board of Education. Collections of measurable objectives in many subject areas are also available from such sources as Instructional Objectives Exchange. Textbook companies also generally have objectives listed in their materials and these may be modified to fit the needs of an individual district. Curriculum guides from neighboring districts are still another useful source.

Balance: Care must be given to avoid an imbalance of objectives. A review of Benjamin Bloom's *Taxonomy of Educational Objectives* will assist the writer in planning for all levels of objectives, from the more literal to the advanced higher levels. In addition, many

subject areas will need to reflect instruction in the affective areas as well as the cognitive areas.

Wording: Writers of objectives should also be reminded to use concrete, measurable verbs where ever possible (e.g., write, compare, say aloud, list versus enjoy, appreciate, know or understand. It should also be noted that objectives may be written for a whole group or program as well as for individuals. (e.g., seventy-five percent of the sixth graders will score seventy-five percent or better on a particular item).

3. Review Research to Ascertain Compatibility with Objectives.

The review of relevant and current research should logically be done as the subject area objectives are being reviewed or designed. The purpose of this step in the test development process is to update the knowledge of the committee as they begin to develop tests. An example of the importance of a review of current research can be found in the area of reading.

The focus in reading in the past decade has moved away from discrete word decoding skills to comprehension strategies (i.e. prediction, clarification, summarization). Reading tests, until very recently, have concentrated on assessing students' mastery of specific skills. With the focus in reading having shifted to comprehension as demonstrated by the competent reader, it is **now necessary** to design tests that are able to assess students' comprehension strategies especially as used in extended selections of connected text. The review of current research and theory in reading, as in other subject areas, should help set the stage as the committee prepares to write tests.

There are a number of sources for current research in all subject areas. The American Educational Research Association acts as a clearing house for many disciplines. The Association for Supervision and Curriculum Development is another association with a goal of disseminating current research and thought on a number of topics. Journals from specific disciplines (math, science, social studies...) are still another source of reports on current research. Knowledgeable school staff, as well as

university and Illinois State Board of Education staff, can also be helpful as the committee completes this phase of the test development process.

The committee will now be ready to review or write the first draft of objectives based on current research and theory.

4. Larger Group Review.

Following the review or development of an initial set of objectives it is advisable to involve a larger group in the review process. Not only does a second group of reviewers present a fresh perspective but it encourages a greater number of people to feel ownership in the process. These additional reviewers might represent a wide sampling of the staff, the remainder of a department or grade level or perhaps even a mixture of both teachers and parents. The purposes of this temporary group would be to:

- Find omissions,
- Locate errors and/or inaccuracies,
- Enlarge the staff commitment to the assessment system,
- Help with the bias review (see Step 6).

5. Develop Test Items and Standards/Criteria.

Determine Best Test: As stated earlier, there are many ways for assessing pupil performance, including but not necessarily limited to criterion reference tests, performance observation check lists and writing samples. Generally, a variety of methods will provide the most complete assessment of the students and programs. A standardized test may be appropriate for measuring grammar skills but a writing sample will more adequately measure a student's ability to translate thoughts to print. A criterion referenced test would be fine for assessing knowledge of short and long vowels but a performance check list would be more effective in discovering a student's ability to speak clearly in an organized manner.

Determine Standards: After determining the type of assessment, the test designers must decide the standard acceptability for

each segment or component of the test. A performance scale assessment may be expressed in binary terms: *pass or fail, happen or not happen, do or don't.*

Criterion referenced tests will result in raw scores or numbers of right and wrong. Levels of success can be established as either a percent or number correct. For example, three out of four on a subtest or seventy-eight percent correct or each sub area.

Writing samples may be scored in a variety of ways. The Illinois State Board of Education, however, is using a model called holistic/analytic scoring for the state writing assessment and districts may want to consider adopting or adapting this model. As mentioned earlier, the Educational Service Centers are prepared to help train people to use this scoring system process.

Test Question Format: Machine scored answer sheets work well with criterion referenced tests in which A-B-C-D answers or true/false responses are required. Generally, however, only single answers to an item may be used with the machine scoring.

Check lists for observation of performance can be designed to meet the specific objectives or items being assessed. Forms should be clear, have large spaces for teacher response, and require as little writing as possible. (See sample of sixth grade listening in Appendix E).

Scoring: There are several decisions regarding scoring that need to be considered. The teacher may score his/her own tests and record the scores. Machine scored tests can be completed by either the teacher or a secretary.

A corps of professionals may be trained to score a particularly difficult type of test or to score a random sampling after the regular staff has scored the tests. A writing sample scored holistically/analytically is an excellent example of where this technique might be appropriate.

6. Conduct Bias Review of all Material, and Amend.

The following comments draw heavily on the booklet *Bias Issues in Test Development* written by the National Evaluation Systems, Inc. in 1987. Readers are strongly urged to read this section as an introduction to the topic of bias review and then to request a copy of the above booklet from the Illinois State Board of Education. *Bias Issues in Test Development* was designed to alert those involved in the test development process to several major areas in which bias in tests may be found. The booklet provides concrete examples of bias (primarily in the areas related to minority groups, women, and people with disabilities) and suggests alternatives.

The test development process should be designed to create test items that are fair to groups against whom there historically has been discrimination (e.g., racial and/or ethnic minorities, women, older persons, persons with disabilities). It is important to address issues of bias at the **beginning** stages of test development, as well as, at the **end** of process of test development. All those involved in the process need to be guided by an awareness of bias concerns.

There are at least three categories of test bias-stereotyping, representational fairness and content inclusiveness. Stereotyping is an image formed by ascribing certain characteristics to all members of a group. A stereotype describes the individual wholly in group terms, making recognition of individuality difficult. Discrimination and bias may also occur either through the underrepresentation of females, minorities, and persons with disabilities in test materials or through their overrepresentation. The former obviously fails to acknowledge the achievements of such groups while the later tends to create stereotypes that do not reflect reality for many members of a group.

Finally, the question of fairness in test materials also involves the issue of content inclusiveness (i.e., the contributions, issues, and concerns of racial and ethnic minorities, females, and persons with disabilities). The issue of content inclusiveness is relevant to any subject area when cultural and/or historical topics are included in the broad definition of the area. History is an area in which the contributions of minorities, for example, have often been overlooked.

People developing tests also need to be knowledgeable of the guidelines for fair representation of minority groups. In presenting the

background of individuals, there can not be suggestions that any racial, ethnic, or religious group is more or less worthy than any other. Test items need to represent various minority groups in all areas of life and culture. Whenever possible, it needs to be made clear that every group has its share of writers, artists, thinkers, scientists and leaders.

Committee members also should be aware of the issues involved in the equitable treatment of the sexes. Girls and boys and women and men should be shown with education and training in all fields and at all levels. Every attempt should be made to eliminate job stereotypes for both women and men. No job should be sex-typed. Furthermore, it should not be implied that some interests and activities are intrinsically masculine or feminine. Finally, whenever the activities and experiences of men are discussed, attempts should be made to include the achievements of women.

Legislation has certainly focused attention on the need for fair and adequate representation of people with disabilities. This legislation has resulted in increasing participation of students with disabilities in our schools. Discrimination against people with disabilities does still exist, however, and this, too, will be reflected in test materials. Language that discriminates against people with disabilities should be recognized and corrected. It dehumanizes people and separates them from the rest of society.

7. Pilot Test.

It is now time to pilot the test using a small number of students. Many times the teachers on the committee will volunteer to pilot the test on a class or small group of their own students. While this practice is desirable, there is also merit in having a teacher(s) not on the committee pilot the test.

There are a number of purposes for piloting the recently developed test. The test format and directions need to be clearly understood by both the teacher and the students. Secondly, the content of the test items needs to be explicit and precise, as well as, appropriate for the students being tested. Certainly items that many miss may flag a badly written item or an item the students are not prepared to answer. Finally, the teacher piloting the test will need to record the amount of time it takes to administer the test, as the length of the test may need to be adjusted.

The results of the pilot testing need to be reviewed with the original committee, as well as, the larger group that reviewed the objectives (step 4). Revisions are made at this time and plans for retesting (if appropriate) are made.

8. Test Administration Policy for Consistent Results.

To assure consistent results among various classrooms and during repeated administrations, directions need to be developed with great care. Consideration should be given to the following:

Time allocations: Will the students have a specific amount of time in which to take the test or may they take as long as desired or until about 80-85% of the class has finished?

Oral and written directions: Directions need to be clear, exact,and consistent from room to room. Examples of the type of questions test makers should ask follow. May the students ask questions of the teachers? May directions be repeated? Is a special classroom arrangement necessary? Are written directions to be read aloud by the teacher? Should the teacher read aloud the questions for children having reading difficulties? May a calculator be used? Exactly what equipment or arrangements are necessary for performance type activities?

Put in writing for the examiners as many answers to specific procedural questions as possible. In addition, whenever possible, try out written directions for the students on actual classrooms or students during the developmental stages of the test.

Physical arrangements: Will the administration of this test work best in a regular classroom situation or can large groups be assembled in a school cafeteria? Is a sound proof area necessary? Will the bell system of the school need to be adjusted?

Date and time of administering the tests: By what date must a specific assessment instrument be completed? May teachers decide when to give the test or is the test date fixed? Must students have completed all material or is the assessment given regardless of completion of the text? A discussion of these and other related questions is essential.

Scoring of the assessment: There are several decisions to make. Will the teachers score the tests or will they be machine scored? Is the scoring subjective enough that an outside source (other teachers) should score a random sample and compare the results? (e.g., the writing sample). Should an independent examiner review the entire process?

9. Conduct Second Bias Review.

The second bias review needs to follow the same guidelines as suggested in step six of this test development process. Again, the authors of this document refer the reader to *Bias Issues in Test Development.*

10. Establish a Standard Criteria for Passing.

Following the administration of a pilot testing of the material, levels for passing must be established. They might include:

-Levels of performance on a listening test or physical activities measure.(See Appendix D for an example of established level of performance.) For example, the standard might be that a student needs to get 'x' number of items correct or perhaps the criteria would be either 'yes' he can or 'no' she can't.

-A second issue to consider is whether to test the entire group or just a random sample of the population. If a random sample is to be used, it must be determined how many and by what method the sample selection will be made.

-Criterion referenced tests must have a level of success established. Must the student get 75% of the items correct or will 98% be necessary? What percent of an entire group must be successful before you feel the program is working? (e.g., a standard of 78% of the students must get at least 85% of the items correct).

11. Develop Reporting, Remediation Procedures, and Staff Development.

The fourth chapter of this document addresses the issue of school improvement based on assessment results. The chapter describes the process of reporting assessment results to staff, principals, school boards, students and the public.

Chapter 3 of this document describes a specific way of scoring and analyzing the results of criterion referenced tests. These analyses have proven to be invaluable when reporting CRT results to staff, administrators and school boards.

Certain types of assessment may require training of those personnel designated to administer the instruments. An an example, staff would need to be trained to score a standardized direct writing sample. Therefore, arrangements would need to be made for appropriate staff development.

12. Administer, Score, and Analyze Tests.

The tests need now to be administered, scored, and analyzed as per the guidelines established in this process. Accurate records of results need to be assembled in usable form and a plan developed for maintaining these records.

13. Report to Public and Staff.

The reader is referred to Chapter 4 for a discussion of reporting to the public and staff.

SUMMARY

The second chapter of this handbook presents information on three kinds of locally designed tests -- the criterion referenced test, the performance scale and check list and the standardized direct writing sample. A sample of each is included in the Appendix. The chapter, in addition, describes and recommends a particular test development process.

References

American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1985). *Standards for Educational and Psychological Testing.* Available through Illinois State Board Education.

Bloom, Benjamin S. *Taxonomy of Educational Objectives.* New York: David McKay Company, Inc., 1956.

Illinois State Board of Education (1986). *State Goals for Learning and Sample Learning Objectives.*, Springfield, Illinois.

Illinois State Board of Education (1987). *Write On Illinois!* Springfield, Illinois.

Instructional Objectives Exchange. Box 24095, Los Angeles, CA 90024

National Evaluation Systems, Inc. (1987). *Bias Issues in Test Development.*

Ory, J. *Improving Your Test Questions.* Urbana-Champaign, Il; University of Illinois, Office of Instructional and Management Services.

Shrock, Sharon and Bill Coscarelli. "Ringing out the Bell-Shaped Curve". *Data Training,* (1986), 48-51.

ELEMENTARY MATHEMATICS CRT PROGRAM
Grade 4 Test
Springfield Public School District #186
Springfield, Illinois

APPENDIX A

Students will need a metric and standard ruler for this test.

1.) Write the digits that go with these words: fifty-sixth. (IA-1)

a. 65
b. 56
c. 50
d. Not given.

2.) Reduce $\frac{4}{8}$ to lowest terms. (IA-2)

a. $\frac{1}{2}$ c. $\frac{8}{4}$

b. $\frac{2}{4}$ d. Not given.

Add or subtract. (IB-1)

3.) $\begin{array}{r} 4771 \\ -2090 \end{array}$

a. 2721
b. 2681
c. 2691
d. Not given.

4.) $\begin{array}{r} 5679 \\ 3041 \\ +\ 501 \end{array}$

a. 8620
b. 8720
c. 9221
d. Not given.

Multiply or divide. (IB-2, IVB-1, IVC-1)

5.) $\begin{array}{r} 9 \\ \times\ 7 \end{array}$

a. 63
b. 36
c. 54
d. Not given.

6.) 72 divided by 8

a. 7
b. 9
c. 8
d. Not given.

7.) Tom went to the store and bought two oranges at $.05 each and one apple at $.08 each. How much did he spend all together? (IC1-2-3)

a. $.15
b. $.13
c. $.18
d. Not given.

8.) Identify the following number: four hundred twenty-two thousand, one hundred two.

a. 122,402 (ID-1)
b. 422,102
c. 422,120
d. Not given.

9.) Fill in the blank using <, > or =. (ID-2, IVB-2)

8,832 ____ 8,999

a. <
b. >
c. =
d. Not given.

10.) Identify the correct fraction for the following: forty-three and one-half. (ID-3)

a. $\frac{1}{2}$ b. $42\ \frac{1}{2}$

b. $43\ \frac{2}{2}$ d. Not given.

Find the missing number. (IE-1, IG-1, IVB-1)

11.) 2 + 3 = ____ + 2

a. 5 c. 3
b. 2 d. Not given.

12.) 5 x 2 = _____ x 5

a. 10 c. 2
b. 5 d. Not given.

13.) (5 + 6) + ___ = 5 + (6 + 7)

a. 7 c. 13
b. 18 d. Not given.

14.) (7 x 9) x 2 = ___ x (9 x 2)

a. 7 c. 2
b. 9 d. Not given.

APPENDIX B

82. Which of the following is a staff?

a) b) c) d)

83. Which of the following is a time signature?

a) b) c) d)

84. A quarter note looks like this a quarter rest looks like?

a) b) c) d)

85. A whole note looks like this a whole rest looks like?

a) b) c) d)

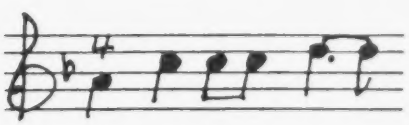
86. An eight note rest looks like

a) b) c) d)

87. A half note looks like this a half note rest looks like?

a) b) c) d)

88. This symbol in the song below is called a?



a) tie

b) phrase line

c) repeat sign

89. This sign means to?

a) play louder b) play faster c) repeat the pattern

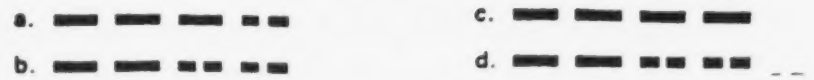
90. Which phrases are alike?

a) a & b

b) b & c

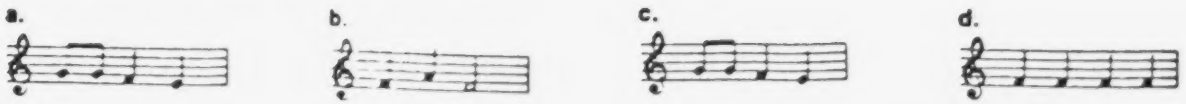
c) a & c

91. Which rhythm lines are the same

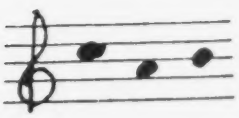


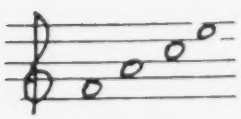
a) a & c b) b & c d) b & d d) a & d

92. Which melody notes are the same?

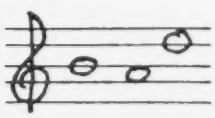


a) a & b b) c & d c) a & c d) b & d

93. The letter name of these notes spell the word.

a) CAB b) TOY c) FACE

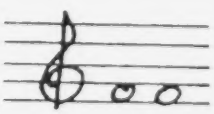
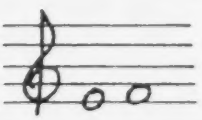
94.  The letter name of these notes spell the word.

a) FACE b) GAME c) EVERY

95.  The letter name of these notes spell the word.

a) BAD b) CAT c) BOY

96. Which pattern spells the word "BE"?

a) b)  c) 

97. This note is

a) F b) E c) J

PERSUASIVE PROMPTS

Teacher Instructions:

Adjust each idea to meet the needs and interests of your students. Share the prompt with the class. Then allow students about twenty minutes to write their original draft. Ideally, teachers would, on subsequent days, have students review their papers with a partner and then write a final copy for your review.

- -

Distribute to class:

BREAKFAST CEREAL

Your mother is going to the grocery store. You want her to buy your favorite cereal. Write an essay convincing your mother why she should buy the cereal you want.

- - Name your cereal and give at least three reasons to support your choice.

- - Support each of your reasons fully with examples and explanations.

- - Explain why this cereal will be good for you.

General Reminders

Take time to plan your essay. Take one position and state it clearly in the beginning. Stay on the topic throughout and be sure to have a conclusion. (FOCUS)

Organize your ideas carefully. (ORGANIZATION)

Explain each reason with sufficient examples and details. (SUPPORT AND ELABORATION)

Check that you have correct sentences, punctuation, and spelling. (CONVENTIONS)

VOCAL MUSIC PERFORMANCE TEST - GRADE 6

1 = Exceptionally proficient in this area
2 = Above average ability
3 = Average ability
4 = Needs improvement
5 = Not Satisfactory

ITEMS TO BE TESTED

1. Given T26 in 5th McMillan, the student Will be able to conduct the 5 songs in the following meters: $\frac{4}{4}$, $\frac{3}{4}$, $\frac{2}{4}$, $\frac{3}{8}$, $\frac{6}{8}$, demonstrating an understanding of "upbeat" and "downbeat", with 80% accuracy.

2. Given listening exercise "What Do You Here 18" in Silver Burdett, Grade 4, the student will select those songs with meters in 2, meters in 3, and meters in 5, with 80% accuracy.

3. Given "What Do You Hear 1", in Silver Burdett Grade 5, the student will select correct rhythm pattern with 80% accuracy.

4. Given p. 70 in McMillan 5, the student will be able to clap the rhythm patterns from this notation with 80% accuracy.

5. Given a simple song the key of F, C, or G, and using the I, IV, and V Chord, the student will be able to accompany the song on the autoharp with 80% accuracy.

6. Given Practice Activity 11 on page CM20 of 6th grade McMillan, the student will correctly identify the 5 answers to the questions on form, with 80% accuracy.

7. Given "What Do You Hear 7" p. 59 of Silver Burdett 4, the student will select the appropriate dynamic markings for the 6 listening examples with 80% accuracy.

8. Given "What Do You Here 4" in Silver Burdett 5, the student will identify by writing the letter of the instrument, the timbre of each instrument as it is heard with 80% accuracy.

9. Given "What Do You Here 9" grade 3, p. 141, the student will correctly select the instrument being played, based upon the tone color heard, with 80% accuracy.

10. Given one line of written notation, (see attached sheet), the student will correctly sight-sing a song from the notation, with 80% accuracy. Suggested songs: Hot Cross Buns p. 26 SB2, Lady Come, Mary Had, Brother John, Sarasponda, America, Jingle bells. Procedure: pass out 1 of the above songs to each student, repeats are fine. Go around the room, one student at a time, asking the student to sing the song on his sheet.

11. Given "What Do You Here 14" in Silver Burdett 3, the student will circle the word "same" or "different", depending on whether the 2 songs being compared are from the same period or not, with 80% accuracy.

Course _____ Instructor _____

Vocal Music Performance Record Grade _____

Rate each student on each area using a scale of 1-5
Each student must demonstrate 80% accuracy overall

1 = Exceptionally proficient in this area
2 = Above average ability
3 = Average ability
4 = Needs improvement
5 = Not satisfactory

STUDENT'S NAME	ITEM 1	ITEM 2	ITEM 3	ITEM 4	ITEM 5	ITEM 6	ITEM 7	ITEM 8	ITEM 9	ITEM 10	ITEM 11	ITEM 12	ITEM 13	ITEM 14	ITEM 15	ITEM 16	ITEM 17	ITEM 18	ITEM 19	ITEM 20	ITEM 21	ITEM 22
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						
_____																						

Goal II. DISTRICT #186 - LISTENING ASSESSMENT CHECKLIST - Grade 6 Page ___ of ___

Name	Generates Questions From Oral Message A1	Draws Conclusions A2	Recalls Significant Details B1	Follows Multi-step Directions B2	Understands Verbal and Non-verbal Cues C1	Identifies Differing Perspective/ Points of View E1	Responds to Speech In Appropriate Manner D1,2 and F1-4

II-25

APPENDIX E

Date ___ School ___
Grade ___ Comments ___
Teacher ___

Key: + Progress
√ Needs Improvement
– No Progress 12-30-87

Goal IV.

DISTRICT #186 - SPEAKING SKILLS CHECKLIST - Grade 6

Name	Speaks Clearly and Expressively A1	Demonstrates Poise A2	Presents Information In Orderly Way B1	Supports Ideas C1	Distinguishes Fact From Opinion C2	Uses Appropriate Verbal and Nonverbal Cues D1	Uses for Appropriate Social Courtesies E2	Recognizes and Uses For Different Purposes, Informing and Entertaining E1,3

II-26

Date _____ School _____

Grade _____ Comments _____

Teacher _____

Key: + Progress

√ Needs Improvement

− No Progress

12-30-87

ITEM ANALYSIS
AND SCORING ASSESSMENT INSTRUMENTS

Introduction

As a result of the Illinois "Reform Package" mandates for education and an increasing interest on the part of educators in assessing student outcomes, many school districts are seeking methods to assess and analyze student academic performance. While norm referenced achievement tests have been used to provide some information relative to students' academic performance, more and more educators have begun to look at criterion referenced measures as an additional method which can be used to provide data based upon local school district expectations.

The Springfield CRT System is a scoring and analysis package for criterion referenced tests. This program processes objective test data from Scantron mark sense forms. The program consists of four major sections:

1. Creation and maintenance of a student information data base
2. Creation and maintenance of a test key and criteria
3. Processing of optical scan forms
4. Analysis of test data and printing of reports.

The Springfield CRT System can be used by any school district in terms of criterion referenced tests but is best suited for schools with less than 1000 students taking an individual test because of necessary clerical time to enter and process the information.

Two versions of the software, DOS 3.3 and PRODOS 8, have been developed and are currently available. The staff of the Springfield Public Schools is presently developing and refining a test generator software package.

Section 1
How To Develop Your System

Upon purchase of the necessary hardware and software, the Springfield CRT System will provide a rather complete system for your use. However, in order to make the best use of the system, a number of decisions must be made by each school or school district in terms of using the system.

Development of a Testing Strategy

One of the first tasks you need to consider is the determination of the roles that will be played by various personnel in your school or district. Specific suggestions concerning roles are made later in this document; however, it is important to consider what will work in your school or school district. We strongly recommend early involvement in the CRT process by all personnel who will be connected with the program.

A second area you will need to address is how your CRT assessment program will interface with existing testing programs you may have in place. We do not view CRT data as providing all the information you may need to properly assess your curriculum. Rather, we view CRT's as providing additional information for your use. While the Springfield CRT System will provide you with information relative to how well your students are performing in comparison to locally-established objectives, we also see a need for norm referenced tests to allow comparison of your students' academic performance to other students in the nation. By planning in advance, you can determine what you want from your complete testing program and avoid unnecessary duplication or the waste of valuable classroom time.

You will also need to consider what you want to assess with the Springfield CRT System. Undoubtedly, you will most likely find the system to be very valuable for providing the data you will need relative to the Illinois State Board of Education's Sample Learning Objectives (SLO's). The Springfield CRT System was designed for use with an objectives-based curriculum and, thus, will provide you with data you can use to determine how well your students are functioning given specific objectives and criteria. Therefore, you need to consider what grades and/or subject areas you want to assess with this system. We would recommend that you phase in the Springfield CRT System, beginning with a few grade levels and subject areas.

We would not recommend starting with an overly-ambitious program of assessing all grades and all subjects.

We also recommend that early in the planning phase users of the Springfield CRT System develop timelines for testing. Timelines, of course, should take into consideration necessary reports which must be filed with the Office of the Illinois State Board of Education but also should reflect consideration of the interfacing with other testing programs as well as the local school or district calendar. Furthermore, your local timeline must be constructed based upon how you want to use the system and the reports generated by the system.

The Springfield CRT System is a rather flexible system which you can use in a variety of ways in your school or district. You will need to review the section in this document which discusses in detail the use of results and determine what information you want the system to provide to you. This decision is critical in the planning phase.

Development of Criterion Referenced Tests

Criterion Referenced Tests by definition are tests designed to measure student performance relative to an established objective(s). The Springfield CRT System is designed to process and report CRT data based upon objective types of tests. To be fully effective, test items must be identified with an objective. Additionally, the user can determine what criteria or level of performance is acceptable for each objective or unit tested.

If you already have in place an objectives-based curriculum and CRT's to assess performance, the Springfield CRT System is ready made for your use. By using this system, you will be able to provide information and feedback not only to teachers, but also parents, members of the Board of Education, and other interested citizens.

If you have an objectives-based curriculum but are not presently using CRT's to assess student performance relative to your curriculum, you can choose to proceed by use of one or more options available to you.

OPTION 1 - Locally Constructed Tests

Criterion Referenced Tests which match your objectives can be written by members of your teaching staff. By using this method, you assure a rather close match between assessment (the tests) and objectives (what the teachers are teaching). This process also typically leads to greater "ownership" in the CRT program by teachers. The negative to this option depends upon how well the actual test items are written. The reliability and validity of test items and, thus, each test, rest upon the ability of teachers to write good test items. Our experience indicates that some of our best teachers may not always be good test writers.

OPTION 2 - Published Tests

More and more publishers are beginning to produce criterion referenced tests or test banks. Use of published tests or test items certainly reduces the amount of staff time required to put a CRT program in place. Furthermore, use of published tests or test items can reduce the risk of "bad" test items dependent upon the care taken in construction by the publisher. The negative side of using published tests or test items centers around the match to your local curriculum. If the test or test items selected for use do not match your local curriculum, you will not assess your curriculum. Furthermore, teachers may not feel as much "ownership" in a CRT program which uses published tests.

OPTION 3 - A Combination Program

You may want to consider using some locally developed test or test items and some published tests or test items in combination. Perhaps you will find that a locally-developed test would best serve a particular high school class while a published test would be suitable for assessing third grade math. Obviously, time and expenses have to be considered. However, we strongly encourage you to consider teacher "ownership" in the program as well as the match between your local curriculum and the criterion referenced test.

Scoring the Criterion Referenced Tests

You will need to make the following decisions during the planning phase relative to scoring your tests:

1. Who will score the tests? - Refer to the section concerning roles in this document for suggestions.
2. When will the tests be scored? - This should be part of your testing timelines.
3. Where will the tests be scored? - You will obviously need some space to actually score the tests.

Section 2
How To Run the Springfield CRT System

System Operation

If you are using an Apple IIe, be sure the [Caps Lock] key is depressed. Be sure to depress the [Return] key after each input action. You will need two disks - one the CRT Program disk and the other the data diskette. The data diskette should be an initialized diskette (ProDOS 8). Information on how to initialize a diskette may be found in the Apple computer owner's manual or any reference text. Please note that that this version will utilize either a 5.25" floppy diskette, a 3.5" 800K hard diskette, or a hard disk. The CRT program is supplied on a 5.25" diskette but may be copied to a hard disk for ease of use. The CRT program will then create and maintain the data files required for the system operation.

When the CRT program disk is booted (CRT program diskette in Drive 1 and the data diskette in any other Drive), a screen appears, describing how to choose files, followed by a menu screen, indicating the student master file and the student Scantron scores file as well as information about the location of disk drives, printer, and Scantron. As a general rule, the only information that would ever need to be changed would be the data volume name and the two file names. A file name will have to be indicated for the student master file and Scantron scores file as well as the data volume. Information at the bottom of the screen indicates how to make changes. [Escape] goes on to the next screen with no changes.

The next menu is the main menu. All modules of the program start from this menu.

```
MAIN MENU:

 0    QUIT
 1  CHANGE VOLUME NAME, FILE NAME, ETC
 2   CREATE/UPDATE STUDENT INFORMATION FILE
 3   VIEW DATA IN STUDENT INFORMATION
           F I L E
 4   CREATE/UPDATE TEST KEY AND CRITERIA
 5   PROCESS SCANTRON SHEETS
 6   PROCESS TEST DATA AND PRINT RESULTS
 7   PRINT STUDENT NAMES AND ID NUMBERS
       ON CONTINUOUS FEED SCANTRON FORMS

ENTER YOUR CHOICE:
```

Choice 1

Returns to the setup menu previously discussed and would be used when changing from one test to another.

Choice 2

This module allows creation and manipulation of the student information file (the student master file).

When Choice 2 is selected, the following menu is displayed on the screen. Make sure that the appropriate data volume is on line in one of the available online disk drives contains an initialized disk if you are creating this file.

MENU:

0 QUIT
1 CHANGE STUDENT NAME, SCHOOL, OR TEACHER
2 ADD STUDENT TO MASTER FILE
3 ADD WHOLE CLASS TO MASTER FILE
4 DELETE STUDENT FROM MASTER FILE

ENTER YOUR CHOICE (0-4)

Enter the student name (last name first) with a space (no comma) between the last name and the first name. The ID number is required to identify students' information when their individual test answer sheets are read by the Scantron optical scanner. Fifteen (15) characters are allowed for the name. Enter the student ID number and the number of the school attended. Each teacher can be assigned a two-digit number with the third digit for different classes taught by that teacher. Successive students for the same teacher and school may be entered without retyping 3 and 4, just hitting the [Return] key. After information has been saved to the data volume, it can only be changed by option 1 on the CREATE/UPDATE STUDENT INFORMATION FILE. The other options allow maintenance of the student master file. Once this file has been created, it may be used to score any number of tests drawing from this student population. This file may be copied to additional data diskettes to facilitate the scoring of additional tests administered to the same group of students.

Choice 3

PRINT INFORMATION ON STUDENTS

Selection of choice 3 allows the user to display student information from the student master file for any class, school or the entire file. This information may be displayed on the screen or sent to the printer.

```
PRINT INFORMATION ON STUDENTS -

1  IN ONE CLASS
2  IN THE WHOLE SCHOOL
3  IN THE WHOLE FILE

YOUR SELECTION?  (0=QUIT)
```

Choice 4

TEST KEY/CRITERIA/MINIMUM COMPETENCY

This program allows you to enter or manipulate the key and criteria for the Scantron data in the following file:

Selection of Choice 3 produces the following submenu on the screen:

```
MENU
 0    QUIT
 1   ENTER TEST KEY
 2   CORRECT TEST KEY
 3   ADD CRITERIA NAME(S)/SUBTEST  QUEST-
       ION NUMBERS/MINIMUM COMPETENCY
 4   DELETE A CRITERION NAME AND SUBTEST
       QUESTION NUMBERS
 5   CHANGE QUESTION INCLUDED IN ONE
       SUBTEST TO ANOTHER SUBTEST
 6   CHANGE SUBTEST MINIMUM COMPETENCY
 7   LIST CRITERION NAME/SUBTEST QUEST-
       ION NUMBERS/MINIMUM COMPETENCY

ENTER CHOICE:
```

Choice 1 allows the user to input the test key into the file. The method is self-explanatory. Choice 2 allows change of the previously-entered test key. Choices 3, 4, 5, 6 and 7 refer to the creation of subtests and the assignment of each question in the key to one of the subtests. Before scoring any tests, options 1 and 3 must be completed. A maximum of 200 questions may be included in the key and up to 16 subtests may be identified. Upon completion of activities from this submenu, Option 0 returns to the main menu.

Choice 5
Scantron Processing

Selection of Choice 5 requires that both the student master file and the Scantron scores file must have been created and they must both be on the data disk before this option may be executed. If the files are correct and the Scantron is powered up, then the following screen is displayed:

BE SURE SCANTRON IS SET TO READ-12

PRESS ANY KEY TO BEGIN AND RUN A DUMMY SCANTRON SHEET THROUGH THE SCANTRON.

AFTER ALL SHEETS HAVE BEEN RUN THROUGH, PRESS REKEY TO END

Press any key and the feed form light should light and the Scantron will buzz. Any form (including a valid test form) can be run through as the dummy form. Forms may be run through as quickly as the machine is able to process the data. The computer will indicate any problems with each form, and the forms may be corrected and run through again or all bad forms may be rerun at one time. Some problems may require the dummy form be run through again in order to clear the system. If the

form is successfully processed, the information will be stored on the data diskette. Please refer to the manual for your scanner if consistent problems are encountered.

When finished, a report of errors will be shown and an opportunity to correct errors will be given either by rescanning or manually correction from the keyboard.

Choice 6

This option analyzes the data stored in the Scantron test file. The correct files must be specified for this option.
The following submenu is then displayed:

```
PLEASE CHOOSE FROM BELOW:

FOR A SINGLE CLASS
 1 COMPLETE REPORT
 2  REPORT WITHOUT STUDENT MATRIX
 3  SUMMARY OF SUBTEST RESULTS
FOR THE WHOLE SCHOOL
 4  COMPLETE REPORT
 5  REPORT WITHOUT INDIVIDUAL STUDENT SCORES
 6  SUMMARY OF SUBTEST RESULTS
FOR THE ENTIRE DISTRICT
  7  COMPLETE REPORT
 8  SUMMARY OF SUBTEST RESULTS

YOUR CHOICE?  (1-8)
```

Selection of one of the report options then produces the following display:

```
ENTER TEACHER CODE

ENTER TITLES (MAX 3 LINES) OR RETURN
ONLY TO QUIT
. . . . . .
TEST
. . . . . .

NUMBER OF                NUMBER OF
RECORDS                  STUDENTS
CHECKED                  READ
000000                     000000
```

The single class complete report option is the only one that produces the student matrix. See the appendices for samples of the various reports. Several classes taking the same test may be combined in the same Scantron file or condensed from several files on several diskettes.

Choice 7

Selection of this option produces the usual request for selection of a student file and then requests the slot number for the printer being used to preslug the test forms.

```
MENU:

0  RETURN TO MAIN MENU
1  ADJUST POSITION OF FORMS IN PRINTER
2  MAKE A TEST RUN (PRINT ONE FORM)
3  BEGIN PRINTING FORMS


YOUR CHOICE (0-3)
```

The menu items for this choice are completed in order of appearance from 1 to 4. They involve adjusting the position of the forms in the printer until they are located in the correct position. This step is critical to successful performance in the scanner. Some larger operations will use an extra printer in the system for forms only to avoid constantly changing paper and adjusting forms. Forms printed by the option are more reliable when scored than forms filled in by students or teachers at the test site.

Backup of Data

Diskettes used by microcomputers are generally very reliable and trouble free but fail without warning. Therefore, backup copies of all data (especially student master files) should be made and kept up to date. A program to duplicate diskettes may be found on the ProDOS System Utilities diskette. No copies should be made of the CRT program diskette.

Section 3
How To Use The Results

The results which you will have available depend upon the testing strategy you have selected. The following discussion will include interpretation of results making full use of all options available. You may decide that you do not want some of the data generated.

Pretest

In terms of the pretest itself, you may want to use the same test you plan to use for the posttest. We have used this strategy and have found little, if any, contamination of posttest results even in our nine-weeks' courses. A better strategy may well be to have available alternate forms of the same test. This would address any concerns for contamination of posttest results. In our estimation, the ideal strategy would be to have an item bank of questions available for each test you plan to give. You can then randomly select questions from the item bank and, thus, customize each test you administer.

If you decide to include a pretest as part of your testing strategy, you will be able to:

A. Identify units or objectives which may require additional attention and/or time

B. Identify students who may require additional assistance.

Identification of units or objectives which may require additional attention and/or time can be valuable information for the teacher. Essentially, what you are measuring is what knowledge your students bring to a specific course or subject area. By giving the pretest during the

first week of the class, you will be able to determine in what areas within the course or subject area your students have strengths or weaknesses which need to be addressed.

Having administered the pretest and scored the results, you can interpret the results as follows:

1. Find the section of the printout entitled SUMMARY OF SUBTEST RESULTS (see Appendix 5 for an example). Using this page, you will be supplied with information concerning the number of students who completed the subtest (unit or objective), the number of students who passed each subtest at the criterion level you established, and the resulting percentage of students who passed each subtest. You will want to examine this information to determine what subtests have the lowest scores and, thus, may require more attention or time. This process is somewhat subjective depending in part on the importance which is placed on specific units or objectives.

2. More detailed information can be obtained by examining the individual SUBTEST RESULTS (see Appendix 3 for an example). The pages of your printout which give you the individual SUBTEST RESULTS supply you with the subtest performance of individual students (raw score, maximum score, and resulting percentage correct), the range of scores, the highest score, the lowest score, the mean, the median, the standard deviation, the error of the mean, and a graph showing the distribution of scores. Using this information, you can determine the extent of the deficiency and, thus, more accurately determine what should be done to address the problem.

3. If more information is desired, you can make use of the STUDENT RESPONSE MATRIX (see Appendix 1 for an example). The STUDENT RESPONSE MATRIX provides you with information concerning the performance of each student on each question of the test. By selecting the questions which make up the subtest you are concerned with, you can examine the specific questions and student responses to determine specifically which test questions caused problems for the students on the subtest being considered.

Identification of students possibly requiring additional assistance can be valuable for teachers, particularly in a course which may be rather short in length. One note of caution, the pretest is only one measure of a student's performance. Other information needs to be considered before final instructional plans are made for an individual student. The information provided by the pretest MAY indicate that an individual student might require additional assistance.

To interpret the results on the printout in terms of students possibly requiring assistance, you would do the following:

1. Complete step three above. By quickly scanning the responses from each student, you can rather quickly determine which student had the greatest problem with the pretest. If a student supplied the wrong answers to many questions, this would be a reason to investigate further.

2. If additional information is desired, you can complete step 2 above and identify specific unit(s) which caused problems for the student.

Once again, please remember that this is only one measure of a student's performance and should be used only as one indicator of possible problems.

Posttest

In terms of the posttest itself, please recall the previous discussion concerning the development of test items and the tests. You may also want to refer to the pretest portion of this section in terms of using pre/posttest design.

By using a posttest in conjunction with the Springfield CRT System, you will be able to:

A. Conduct an item analysis
B. Identify units or objectives requiring possible modification
C. Conduct program assessment
D. Conduct student assessment.

Item analysis will enable you to identify bad test questions, bad foils, or incorrect answers on the test key. This analysis can be conducted using two sections of the printout (see Appendix 1 and Appendix 2 for examples). To conduct an item analysis using the STUDENT RESPONSE MATRIX (Appendix 1), you simply look at each question individually, trying to find questions where large numbers of students responded incorrectly to the question. If you find such a question, you need to:

1. Determine if the correct answer was entered in the test key.
2. Evaluate the question - Was it clearly stated? Are the foils appropriate? Are the available answers clearly stated?
3. Determine if the question was addressed through instruction and under what conditions.

The RESPONSE MATRIX (Appendix 2) should also be used to conduct an item analysis. This matrix is used by examining each question individually

once again. The number of students who responded for each choice available is listed across from the question number and under the choices available. The correct answer is identified by locating the answer bracketed by the two asterisks. Once again, if you find a question which many students have answered incorrectly, you will need to follow the same procedure that was just discussed in terms of the STUDENT RESPONSE MATRIX.

Identification of units or objectives requiring possible modification can be accomplished by use of the SUMMARY OF SUBTEST RESULTS (Appendix 5) and each individual SUBTEST RESULTS sections (Appendix 3) of your printout. By examining the results from all of the subtests, you can rather quickly identify units or objectives which posed problems for your students. These problem areas can then be examined in detail by looking at the individual SUBTEST RESULTS to determine the degree to which the problem needs to be addressed. Individual questions can be examined by use of the STUDENT RESPONSE MATRIX.

Program assessment can be accomplished in a number of ways, depending upon what you want to assess. If you want to determine in general terms the success of your students on the posttest, you will find the SUMMARY OF STUDENT RESULTS (Appendix 4) to be helpful. If you want to determine specific levels of students' performance on a unit or objective, you would make use of SUMMARY OF SUBTEST RESULTS (Appendix 5) and the individual SUBTEST RESULTS (Appendix 3) that you may be interested in. The data contained in these reports should provide almost all the information you would need.

Student assessment can also be addressed to a degree by use of the Springfield CRT System. Particularly in courses which make use of a final examination, we have made some use of the posttest as part or all of the final examination. The benefit of this method is that you standardize final examinations within a building or district and the teachers are not in a position of constructing their own final. Critics of this method are quick to point out that CRT's should assess only whether students have mastered objectives and are inappropriate for student assessment since CRT's should not reflect the normal curve. If you choose to use the posttest for student assessment, you can refer to the SUMMARY OF STUDENT RESULTS (Appendix 4) where raw scores and percentages are reported.

Combination of Pretest and Posttest

By using the data from both pretest and posttest, you can do the following:

1. Conduct program assessment by use of a statistical test
2. Graph results on the pre and posttests.

An additional method of program assessment would be to statistically assess the difference between the results of the pretest and posttest. Use of a test like the t-test of dependent samples can supply you with information relative to this comparison. In essence, what you are asking is, "Is there a statistically significant difference between students' performance on the pretest and the posttest?" The answer to this question can be very interesting as well as powerful in terms of teachers, administrators, the Board of Education, and interested citizens. With the advent of a number of statistics packages for the microcomputer,

the actual calculations involved in the tests are no longer overwhelming. While a statistics package is not included with this program, you may want to consider the purchase of one if you plan to conduct this type of program assessment.

We have found some benefit in graphing the results of students' pretest and posttest results. By overlaying the results from these two tests on the same graph, it is sometimes easier to actually see results. This visual evidence can provide the basis for some very interesting discussions. Once again, there are some microcomputer graph programs available which can make this task very easy to accomplish. While a graph program is not included in this package, you may want to consider the purchase of one.

Section 4
How To Cope With Problems

Even though the Springfield CRT System is essentially trouble free, problems with the system can still develop. We strongly suggest that one or two people be designated to operate the system. By doing this, confusion among those using the system will be reduced and incorrect data entered into the system can be avoided; thus, eliminating many problems before they start.

Based upon our use of the system, listed below are some of the common problems you may experience as well as suggestions for avoiding the problems.

1. Problem: Incorrect ID numbers on the Scantron answer sheets.
   Suggested Solution: Use preslugged forms or check each answer sheet prior to scoring.
2. Problem: Improperly marked answer sheets.
   Suggested Solution: Make sure students use a No. 2 pencil and erase completely. Furthermore, instruct students to select only one answer.
3. Problem: Erasure debris clogging the scanner.
   Suggested Solution: Make sure all sheets are clean prior to scoring through the Scantron.
4. Problem: Diskette failure.
   Suggested Solution: Make sure to have back-up copies for everything

5. Problem: If all else has failed to solve a problem.
   Suggested Solution: Contact Mr. Mike Holinga, Microcomputer Coordinator, Springfield Public Schools, 217/525-3015.

Section 5
Tentative Timeline

In addition to the timeline you will establish for your testing program, it is necessary to take some other time-related items into consideration. First of all, consider what time you will need to allow for purchase and installation of your hardware. Allow sufficient time to provide for possible delays in shipping or installation. Secondly, allow for training time in order that the personnel designated to operate the system can become familiar with the system prior to formal implementation. While we will be offering formal training sessions, we would also suggest an additional day or two, perhaps used in parts, for your personnel to become comfortably familiar with the Springfield CRT System.

In terms of operational timelines, once your system operator has become familiar with the system, and assuming a class of 30 students, you can anticipate the following time requirements:

1. 15 minutes to input names and ID's
2. 15 minutes to input the answer key and criteria
3. 15 - 20 minutes to scan the answer sheets and print the report

It is important to note that items 1 and 2 above will only have to be done once for each class tested regardless of whether you use a posttest only design or a pre/posttest design. Item 2 above will only have to be done once for all classes taking the same test unless you decide to change the test, some test items, or answer choices.

Section 6
Roles of Various People Associated With The System

In this section, we will suggest various roles for individuals associated in some manner with the use of the Springfield CRT system or your CRT program. Obviously, these are only suggestions; but we make them to you based upon our experiences involving the use of this system as well as our CRT program. We hope you will find these suggestions useful as you initiate your program.

Administrators

Use of the Springfield CRT System and the processes we have suggested to you can truly put administrators at the center of that often talked about concept of "instructional leadership". With the data you will have available to you, administrators at all levels should have a much better grasp of what is happening instructionally, including areas of strengths and weaknesses within the instructional program. Furthermore, having this data at hand can put administrators in a position of working with teachers to develop ways to maintain the strengths and address the weaknesses. An administrator has several roles to play in connection with the use of the Springfield CRT System but all are related to true instructional leadership.

Planning - As is the case with most programs, administrators must take the lead in terms of planning. This involves a number of decisions relative to the use of the Springfield CRT System. These decisions range from "Shall we use this system?" to questions of who, when, where, and under what conditions. We strongly encourage the use of a broad base of people early in the planning stage in order to encourage "ownership" as well as provide for smooth operations.

Budgeting - Assuming you determine a need for our system, an administrator will undoubtedly have to secure and allocate financial resources which can be used to purchase and support the CRT program.

Organizing - An administrator should take responsibility for the organizing associated with the use of the program. This role responsibility requires a person to organize the necessary resources required for implementation of the program. Specifically, this requires

acquisition of the necessary hardware and software, training of personnel, and the refinement of plans.

Implementing - An administrator should take responsibility for the implementation of the program. This will require some time to make sure tests are administered based upon the schedule established and are scored and returned within the time frames established.

Evaluating - Administrators should not only be involved in the assessments provided by the Springfield CRT System but also involved in evaluating the operation of the system and the CRT program. Summative and formative evaluations are both useful in this effort.

Teachers

Teachers play a critical role in the Springfield CRT System as they should. Teachers need to be involved early in the planning stage as well as involved in all other stages of use of the system. Teachers can and will provide valuable inputs and in return will receive valuable feedback.

Planning - Teachers can provide useful information concerning the establishment of a testing strategy, development of tests, and development of timelines. This is all information you will need to make good sound decisions.

Organizing - Teachers can assist in organizing by providing a sounding board for administrators as well as providing assistance in spreading the word and selling the program.

Implementing - The teachers' role is critical in terms or implementation. Obviously, without the teachers' cooperation, a testing strategy and timeline can be completely destroyed. That is why early involvement is so very important.

Evaluation - The teachers, like the administrators, will undoubtedly be concerned with the assessments produced by the system. These assessments will provide timely feedback to the teachers who can thus make better decisions concerning their students and lesson plans. The teachers should also be involved in evaluation of the operations of the system and may be in the best position to offer suggestions for improvements.

Clerical

We recommend that one or two of your clerical staff be designated to actually operate the Springfield CRT System. Clerical staff members who are familiar with the system can prove to be great assets to proper and expedient use of the system.

Planning - We recommend that clerical staff members who will be using the system be involved in the early planning stages. Following training and some initial period of familiarization, clerical staff members can provide great assistance in terms of timelines and your testing strategy.

Organizing - Clerical staff members can assist in organizing by serving as a sounding board as well as helping with the training.

Implementing - This role of the clerical staff members is also critical. The primary task associated with this role will be responsibility for scoring the tests and printing the reports. This critical role, improperly fulfilled, can doom your program to failure. Early involvement can reduce problems and result in smoother operations.

Evaluating - The role of the clerical staff members is different to a degree than others associated with the system. The clerical staff members should be entirely concerned with the operations of the system and not assessments produced by the system. Clerical staff members can offer keen insights into how to smooth out your operation of the system.

Board of Education

The Board of Education should have roles in the establishment of Springfield CRT System. Involving the Board initially and keeping them informed can reduce problems you may be forced to deal with.

Planning - We would encourage the Superintendent to recommend the establishment of Criterion Referenced Testing Program to the Board of Education. The extent of information supplied to the Board may vary depending upon what you feel is required, but we feel that the Board should approve of the program through the adoption of a policy statement.

Budgeting - The Board of Education will also have to give final approval to the allocation of funds to establish and maintain the CRT program. This role may very well be connected with the Board's adoption of a policy statement establishing the program.

Evaluating - Once your program is up and operational, we would suggest the Board be informed of your progress. This progress can be reported in terms of operation of the program as well as results of the program. This information will be useful to the Board as they evaluate the program and resulting costs incurred.

Section 7
Materials You Will Need

The materials you will need in order to operate the Springfield CRT System are as follows:

Hardware

1.	Apple IIe computer with 128K memory and disk drives (professional system)	$1014.00
2.	80 column printer (Imagewriter II)	486.50
3.	Super serial card	100.00
4.	Scantron data terminal (Model 1200 or equivalent)	1800.00
5.	Blank 5 1/4" diskettes ($1.00 per disk)	as needed
6.	Scantron Forms (#2200 or 2097c) 1000 forms @	50.00
7.	Blank computer paper (1 part or 2 parts) 2600 sheets @	25.00.

These materials are available through many different dealers and are not available through the Springfield Public Schools. Prices are estimates at the time of publication of this document and reflect the Apple educational discount.

Software

The software for the Springfield CRT System is available for purchase from the Springfield Public Schools for the price of $250.00. This software is copyrighted and can only be duplicated for use as a back-up to the originally purchased software.

CHEMISTRY FINAL EXAM

APPENDIX A

STUDENT RESPONSE MATRIX

.....CORRECT RESPONSES ARE INDICATED BY -
.....INCORRECT RESPONSES ARE LISTED
.....AN INVALID RESPONSE IS INDICATED BY *
.....NO RESPONSE IS INDICATED BY +

STUDENT NAME	QUESTION NUMBER																									NUMBER RIGHT	CUM # RIGHT
	0 0 1	0 0 2	0 0 3	0 0 4	0 0 5	0 0 6	0 0 7	0 0 8	0 0 9	0 1 0	0 1 1	0 1 2	0 1 3	0 1 4	0 1 5	0 1 6	0 1 7	0 1 8	0 1 9	0 2 0	0 2 1	0 2 2	0 2 3	0 2 4	0 2 5		
~~AAAAA~~ JOHN	-	-	-	-	-	-	-	-	-	-	-	B	-	-	-	-	-	D	-	C	-	-	-	-	*	21	21
BBBBB JONATHAN	-	-	-	-	-	C	-	-	-	D	-	-	-	-	-	-	-	D	-	-	-	A	-	-	A	20	20
CCCCCCC JOHN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	A	23	23
DDDDD TONY	-	-	-	C	-	-	A	-	-	A	-	C	-	A	-	-	D	C	-	-	-	A	-	A	C	15	15
EEEEE ROBERT	-	D	-	C	-	-	B	-	C	A	-	-	-	-	-	A	B	D	B	C	C	C	-	-	A	12	12
FFFF THOMAS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	B	-	-	-	24	24
GGGGGG SHAVONNE	A	-	-	-	-	B	-	A	-	-	-	-	B	-	-	-	B	D	-	-	A	-	B	A	C	15	15
HHHHH ANDREW	-	-	-	-	A	C	-	-	-	A	-	-	-	-	-	A	A	-	-	-	-	-	-	-	-	20	20
III CHRISTOPHER	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	A	-	C	-	-	-	-	-	-	D	21	21
JJJJ FREDERICK	D	-	-	-	-	B	-	-	-	-	-	C	-	A	-	A	-	C	D	-	A	C	-	-	C	15	15
KKKKK ARSHIA	B	-	D	-	C	-	-	A	-	D	-	-	-	-	-	D	B	C	-	-	A	A	-	-	A	14	14
LLLLLL ERIC	-	-	-	-	-	B	-	-	-	-	-	D	B	-	-	-	-	-	D	-	D	-	-	-	C	19	19
MMMM ROGER	-	-	-	B	-	-	A	-	-	D	-	-	-	-	-	-	-	D	-	-	-	C	-	-	D	19	19
NNNNNN DOUGLAS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	A	-	-	-	-	D	A	-	-	-	21	21
OOOOO OSCAR	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	24	24
PPPP DEIDRE	-	-	-	-	A	-	-	-	-	-	-	B	-	B	-	C	-	D	-	C	-	B	-	D	-		17	17
QQQQ STEVE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	B	-	D	A	-	+	-	21	21
KEY	C	A	B	D	B	A	D	C	D	B	A	A	C	C	C	B	C	B	A	B	B	D	D	C	B			

REPORT ON STUDENT RESULTS

APPENDIX B

RESPONSE MATRIX

QUESTION	A	B	C	D	E	% CORRECT
1	1	1	* 14*	1	0	82
2	* 16*	0	0	1	0	94
3	0	* 16*	0	1	0	94
4	0	1	2	* 14*	0	82
5	2	* 14*	1	0	0	82
6	* 12*	3	2	0	0	71
7	3	1	0	* 13*	0	76
8	3	0	* 14*	0	0	82
9	0	0	1	* 16*	0	94
10	3	* 11*	0	3	0	65
11	* 17*	0	0	0	0	100
12	* 12*	2	2	1	0	71
13	0	2	* 15*	0	0	88
14	3	1	* 13*	0	0	76
15	0	0	* 17*	0	0	100
16	5	* 10*	1	1	0	59
17	1	3	* 12*	1	0	71
18	0	* 6*	5	6	0	35
19	* 13*	2	0	2	0	76
20	0	* 14*	3	0	0	82
21	3	* 10*	1	3	0	59
22	5	2	3	* 7*	0	41
23	0	1	0	* 16*	0	94
24	2	0	* 14*	1	0	82
25	4	* 7*	4	2	0	41
26	1	4	0	* 12*	0	71
27	* 16*	1	0	0	0	94
28	* 14*	2	0	1	0	82
29	2	* 12*	1	2	0	71
30	1	3	7	* 6*	0	35
31	4	0	0	* 13*	0	76
32	0	* 14*	2	1	0	82
33	5	7	0	* 5*	0	29
34	0	0	* 9*	8	0	53
35	0	0	* 17*	0	0	100
36	* 14*	1	1	1	0	82
37	0	2	0	* 15*	0	88
38	1	0	* 16*	0	0	94
39	* 7*	5	0	5	0	41
40	* 9*	4	0	0	0	53
41	1	0	* 15*	1	0	88
42	0	0	0	* 17*	0	100
43	0	1	* 16*	0	0	94
44	0	* 15*	2	0	0	88

SUBTEST RESULTS

A. INTRODUCTORY CONCEPTS

STUDENT NAME	SCORE RAW/MAX		% CORRECT
AAAAA JOHN	18/20		90.00
BBBBB JONATHAN	16/20		80.00
CCCCCCC JOHN	20/20		100.00
DDDDD TONY	16/20		80.00
EEEEE ROBERT	9/20	*	45.00
FFFF THOMAS	20/20		100.00
GGGGGG SHAVONNIE	14/20		70.00
HHHHH ANDREW	13/20	*	65.00
III CHRISTOPHER	18/20		90.00
JJJJ FREDERICK	15/20		75.00
KKKKK ARSHIA	10/20	*	50.00
LLLLLL ERIC	13/20	*	65.00
MMMM ROGER	12/20	*	60.00
NNNNNN DOUGLAS	17/20		85.00
OOOOO OSCAR	19/20		95.00
PPPP DEIDRE	15/20		75.00
QQQQ STEVE	20/20		100.00

*'= DID NOT MEET CRITERION LEVEL

ABSOLUTE RANGE	HIGHEST SCORE	LOWEST SCORE	MEAN	MEDIAN	STANDARD DEVIATION	ERROR OF MEAN.
11	20	9	15.6	16	3.4	.8

DISTRIBUTION OF SCORES

RANGE %	NUMBER	
100	3	IXXX
95 TO 99.99	1	IX
90 TO 94.99	2	IXX
85 TO 89.99	1	IX
80 TO 84.99	2	IXX
75 TO 79.99	2	IXX
70 TO 74.99	1	IX
65 TO 69.99	2	IXX
60 TO 64.99	1	IX
55 TO 59.99	0	I
50 TO 54.99	1	IX
45 TO 49.99	1	IX
40 TO 44.99	0	I
35 TO 39.99	0	I
30 TO 34.99	0	I
25 TO 29.99	0	I
20 TO 24.99	0	I
15 TO 19.99	0	I
10 TO 14.99	0	I
5 TO 9.99	0	I
0 TO 4.99	0	I

70.6% HAVE MET THE CRITERION LEVEL

APPENDIX D

SUMMARY OF STUDENT RESULTS

STUDENT NAME	SCORE RAW/MAX	% CORRECT
AAAAA JOHN	60/80	75.00
BBBBB JONATHAN	68/80	85.00
CCCCCCC JOHN	73/80	91.25
DDDDD TONY	48/80	60.00
EEEEE ROBERT	40/80	50.00
FFFF THOMAS	72/80	90.00
GGGGGG SHAVONNE	48/80	60.00
HHHHH ANDREW	52/80	65.00
III CHRISTOPHER	67/80	83.75
JJJJ FREDERICK	54/80	67.50
KKKKK ARSHIA	45/80	56.25
LLLLLL ERIC	44/80	55.00
MMMM ROGER	58/80	72.50
NNNNNN DOUGLAS	62/80	77.50
OOOOO OSCAR	75/80	93.75
PPPP DEIDRE	43/80	53.75
QQQQ STEVE	62/80	77.50

ABSOLUTE RANGE	HIGHEST SCORE	LOWEST SCORE	MEAN	MEDIAN	STANDARD DEVIATION	ERROR OF MEAN.
35	75	40	57.1	58	11.4	2.8

DISTRIBUTION OF SCORES

RANGE %	NUMBER	
100	0	I
95 TO 99.99	0	I
90 TO 94.99	3	IXXX
85 TO 89.99	1	IX
80 TO 84.99	1	IX
75 TO 79.99	3	IXXX
70 TO 74.99	1	IX
65 TO 69.99	2	IXX
60 TO 64.99	2	IXX
55 TO 59.99	2	IXX
50 TO 54.99	2	IXX
45 TO 49.99	0	I
40 TO 44.99	0	I
35 TO 39.99	0	I
30 TO 34.99	0	I
25 TO 29.99	0	I
20 TO 24.99	0	I
15 TO 19.99	0	I
10 TO 14.99	0	I
5 TO 9.99	0	I
0 TO 4.99	0	I

SUMMARY OF SUBTEST RESULTS

APPENDIX E

SUBTEST DESCRIPTION	# TESTED	PASSED	% PASSED
INTRODUCTORY CONCEPTS	17	12	70.6
PHYSICAL CONCEPTS	17	11	64.7
ATOMIC & MOLECULAR CONCEPTS	17	8	47.1
SOLUTION CONCEPTS	17	10	58.8

USING ASSESSMENT INFORMATION FOR SCHOOL IMPROVEMENT

Introduction

Part of a district's belief system must be that information gathered through assessment procedures is useful in some way to someone. The accountability of schools for student achievement presupposes that assessment is not merely the process of testing children and compiling the data, but must include in it the process of determining whether or not student performance is at the desired level. In this instance assessment must lead directly to evaluation. Evaluation, by our definition, refers to *what one does with the information gathered in an assessment program.*

The use of assessment information for school improvement may be accomplished in various ways. As noted previously in this document, those responsible for its content believe that a school improvement team, developed and focused at the building level, offers one appropriate alternative to managing assessment information. In this chapter somewhat formal suggestions are presented regarding two basic topics. The first relates to the utilization of **improvement teams**, and the second relates to organizing and interpreting assessment data with **school profiles**. School units, when considering various alternatives, may choose a formal procedure or one less formal than the one described here. The decision about the structure and formality of school improvement teams is likely to be related to the size of the school district and the number of schools in it. The important concern for school personnel to keep in mind when constructing school improvement teams is devising a systematic approach that addresses both building needs and district needs, rather than approaching any one model as **the best** alternative.

Program Evaluation for School Improvement

School personnel and members of boards of education will assume various roles and responsibilities in the process of evaluating school programs, regardless of the model a school unit develops for using assessment information. Some individuals or groups of individuals will

participate directly and formally in program evaluation and school improvement efforts. Others may have an interest in either or both the evaluation or the improvement efforts that grow out of evaluation, but they may not participate in a formal sense. What roles and responsibilities one assumes will probably depend significantly on his/her position in the system, the size and organizational structure of the system, and the degree of participative decision-making planned for and accommodated in that system. The roles and responsibilities of potential participants are described in the following sections.

District-level

At the district level, the useful components of a comprehensive assessment system would most likely be norm referenced test data, criterion referenced test data, and the summary information from check lists, rating scales and writing samples. These types of data most likely would be useful to decision-makers at this level because these individuals are interested in analyzing the performances of groups of students rather than the performances of individual students.

The **board of education** should expect and, indeed, require assessment data. As the policy-makers of a school district, the school board has the responsibility to see that schools are fulfilling their mission -- effectively educating children. By reviewing assessment data the school board will know if schools have been successful. In addition, the school board has the responsibility of providing adequate resources to carry on assessment activities and, where necessary, activities directed toward school improvement. Finally, the school board must communicate to the appropriate publics how well schools are doing based on the results of assessment data.

The **superintendent of schools** uses test data in reporting about student achievement to the board of education. Norm referenced achievement tests given to selected grades to determine how well students score when compared to other students nationally produce data of interest to members of the school board as do criterion referenced test data designed to determine how well students are achieving locally developed learning objectives. Accountability of schools to the public is most visible when data are reported in these

contexts. The superintendent has an interest, then, in making careful use of assessment results.

The **director and members of a district's curriculum department** will utilize assessment data in different ways. The primary mission of those charged with the responsibility of directing curriculum and instruction is to develop and implement programs for the entire district. In the implementation phase of curriculum development, the evaluation component is of great importance. The use of criterion referenced tests developed during the curriculum writing process and administered on a pretest and posttest basis will provide information relative to how well curricular objectives are being met by the newly implemented curriculum. Adjustment, if needed, can then be made based on assessment results.

Curriculum monitoring, another mission of a district's curriculum department, can continue by periodic examination of district criterion referenced test data. By using CRT's and normative data over a period of time, trends may become apparent that will need to be addressed.

The curriculum department may make use of norm referenced test data by explaining results to building principals and interested others. The impetus for using NRT data to set school improvement goals may come from a district's curriculum department. Typically, district-level curriculum specialists should be expected to serve as resources for building-level improvement teams.

Building-level

The **principal**, as an instructional leader, can make excellent use of assessment information. Norm referenced and criterion referenced test data can help the principal monitor instructional programs by providing information about student achievement in a particular school. When district-wide data and school data are compared, the principal is able to make judgments and, ultimately, decisions about the instructional program in his/her school. The principal, with the assistance of a school improvement team, also is in the best position to collect artifacts and analyze grades as part of program evaluation.

Assessment results will serve as an aid in the development of school improvement plans. As previously discussed, a school-based

school improvement procedure known as **profiling** makes use of norm referenced data, criterion referenced data, and the other components of assessment outlined in Chapter 1 in determining what changes need to be made in the instructional program in order to facilitate student achievement. The astute principal will gather this information and present it to staff members in an effort to gain their active support and participation in the school improvement efforts.

Not all assessment results will show a need for change. Program strengths may become evident and serve as a foundation on which the school principal and staff can build a more solid instructional climate.

By using assessment results to identify both strengths and weaknesses in a school's academic program, the principal can help individual teachers identify their personal strengths and weaknesses relative to the school's. Staff development, then, is another school activity which the principal can support with the use of assessment data. We are not advocating the evaluation of teachers based on the results of their students' achievement test scores. In our opinion, such evaluation would be a misuse of those data. However, making teachers aware of assessment results and helping them develop teaching strategies to enhance student achievement is a valid use of assessment information.

Classroom-Level

Without a doubt the most important place where assessment information is used is the classroom. Here, **teachers** make use of criterion referenced tests, check lists, rating scales, journals, writing samples, and other teacher-made tests. Norm referenced tests have less value to classroom teachers than the above mentioned assessment instruments because information gleaned from normative data is more accurate when applied to a group larger than a classroom of students.

Teachers make the best use of assessment information when it is used as a part of diagnostic/prescriptive teaching both with a class and with individual students. Teachers are more likely to welcome assessment practices when they have fully realized that assessment is a part of instruction, not apart from it. To make judgments about what students know or do not know in the absence of information based upon

accurate assessment is not good teaching. Let us briefly examine the ways in which assessment information can be used prescriptively by the teacher.

To meet classroom needs. In certain situations, a teacher will need to tailor his/her teaching strategies to meet the needs of his/her entire class. By using a pretest the teacher can find out very quickly what the class knows or does not know about a topic before instruction begins. Time is saved because the teacher does not need to rely on trial and error to determine where to begin instruction and does not spend time teaching what students already know. Posttest results will enable the teacher to determine what needs to be retaught or reviewed. The posttest should never be thought of as the "culminating activity" of a unit of study. It is the yardstick by which achievement of certain objectives is measured. If the results fall short of the mark, then the teacher must provide additional instruction to enable students to "measure up" or modify lesson plans for future use.

To meet individual needs. The needs of individual students can be met when a teacher uses assessment information prescriptively. By using the pretest/posttest design described above, the teacher can tailor instruction for individual students. This "teach-test-(re)teach" method of instruction leaves little to chance. It enables the classroom teacher to ensure concepts are learned before moving on to others.

Teachers who develop a strong sense of being accountable for what students learn make use of assessment results in their planning for classroom instruction. For them assessment of student performance is an integrated, routine component of the instructional process.

Planning For School Improvement By Profiling

What is Profiling?

A **profile** can be defined as *a method of examining data about a specific school in order to make decisions to improve aspects of the school.* Data collected are organized in a **profile** which enables

decision-makers to examine student outcomes, school practices, demographics, and resources in enough detail to identify specific areas needing improvement. Through the process of examining the pertinent data, inquiry is stimulated as to why certain conditions exist, such as rising or declining test scores or changes in student attitudes.

Profiling can occur at different levels. School districts can examine individual schools or the district as a whole. State education agencies can use profiling to examine the state system. However, effective schools research has shown that school improvement is best accomplished when improvement efforts are based in the individual school. Therefore, emphasis is generally placed on using profiling in the school setting.

Who does the profile?

Essentially, all people who are involved in delivering an educational program should be involved in developing the school profile. The **building principal**, as instructional leader, will be the person charged with the responsibility of organizing the school improvement effort and keeping it energized. S/he will appoint a **school improvement team** comprised of classroom **teachers** chosen because of their interest in and commitment to the improvement process. The chosen teachers should be influential among their peers, because they must help involve all teachers in the decision-making and program improvement implementation. Remember, research has shown that effective schools empower teachers to become involved in planning and implementing school improvement plans.

The school improvement team should include a **district-level official** who will act as a resource during the planning and implementation stages. This individual should have access to research in order to provide the necessary research base for effective school practices. In addition, this individual will provide the resources (time, materials, consultants) needed for planning and implementation.

What is the process?

The process of profiling was developed by the staff of the Northwest Regional Educational Laboratory (NWREL) by synthesizing the

effective schools research. The ten steps of profiling reported by Blum and Butler (1985) which schools can collegially use for setting school improvement goals based on important data include:

1. Introduce the program to all school staff members, form a **school improvement team** to head efforts, and set the stage for improvements.

2. Study the effective schools research base and findings.

3. Profile current levels of student performance (academic achievement, social behavior, and attitude). See Figure 1 which was used by Blum and Butler (1985).

4. Set a goal for improvement. Remember, your project will be more effective if you limit the number of goals.

5. Check current instructional practices.

6. Develop a research-based prescription to meet the improvement goal.

7. Prepare for implementation.

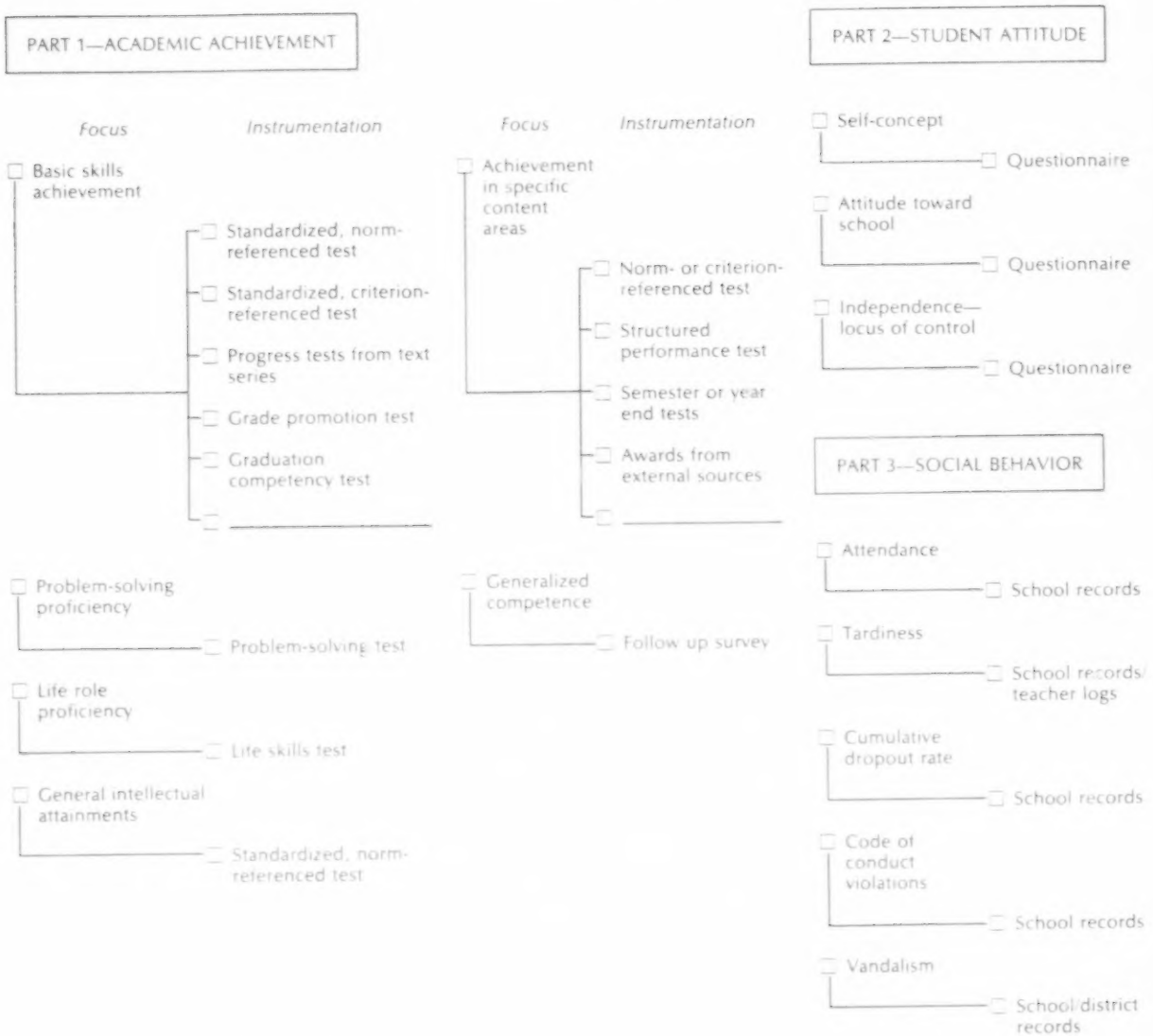
8. Implement the prescription.

9. Monitor implementation.

10. Evaluate progress and renew efforts.

Profiling takes place in steps three and four of the process. In step three, school teams plan for profiling by introducing the process to the staff, collect and analyze data, and prepare the written profile. In step four, the school reviews the data for strengths and weaknesses in student performance and makes recommendations for needed improvements. Goals are set and articulated to the school community.

Figure 1 -- Profile Contents Menu



Note. From "Managing Improvement by Profiling" by R.E. Blum and J.A. Butler, *Educational Leadership*, 1985, *42*, p. 55.

The other steps listed above can be found in other school-based school improvement models; therefore, they will not be discussed here. Steps three and four can be considered the "needs assessment" steps which are necessary in any change or improvement process. In profiling, the needs assessment focuses on student outcomes (student achievement, student attitudes, and student behavior) only. The data

gathered are empirical, not perceptual in nature. Rather than depend on what everyone "thinks" is wrong with the school, reliance should be made on what the data (test scores, attendance figures, discipline data) indicate may need improvement. If, after intervention, your improvement strategy has been successful you will know by reviewing the data.

Let's take a look at steps three and four in some detail.

Step Three -- Data Collection

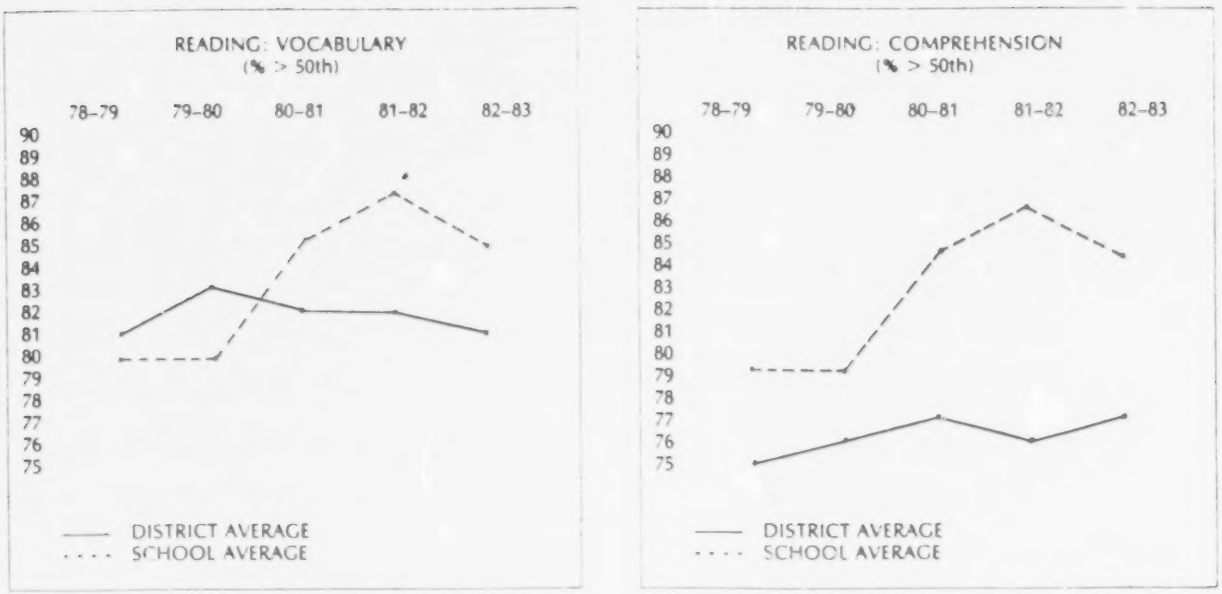
Once the school improvement team has been formed, understands the process being used, and is familiar with the research-base for school improvement, data can be gathered. Figure 1 shows the three areas of student outcomes and various sources for the data needed.

The team decides who will be responsible for collecting pieces of the data. Figure 1 can be used as a check list, checking the boxes as the data are collected or otherwise indicating what is left to be done. Ample time should be given to the collection of these data. The district central office personnel on the team will be very useful as a resource for gathering some of the information.

When data have been gathered, the team will decide how the data will be summarized and displayed. For example, attendance information may best be summarized and displayed graphically. Norm referenced standardized test information will probably best be displayed on a chart or table. The purpose of displaying data appropriately is that a clear, visual presentation of information will aid decision-makers in prioritizing areas needing improvement. Each display should include a short (two or three sentences) narrative summarizing the data. Refer to Figure 2 (Blum & Butler, 1985) which shows a display and narrative for reading achievement.

Team members then collect and organize each data summary and display into a booklet which is disseminated to all staff members for their consideration and evaluation.

Figure 2 -- Display and Summary of Reading Achievement



Narratives

1. The percentage of 7th grade students in the district scoring above the 50th percentile in reading/comprehension skills has increased by 2 percent since 1978–79; and the percentage of those students at this school has increased by 8 percent over the same time period.

2. The percentage of 7th grade students in the district scoring above the 50th percentile in reading/vocabulary skills increased from 1979–80 to 1978–79 by 2 percent and declined between 1979–80 and 1982–83 by 2 percent; the percentage of those students in this school increased overall by 5 percent between 1978–79 and 1982–83.

3. In both skill areas, there was a steady increase in the percentage of students scoring above the 50th percentile over a three-year period between 1979–80 and 1981–82; there was a decline in both skill areas in 1982–83.

4. The percentage of students scoring above the 50th percentile in both skill areas at this school is higher than that percentage in the district.

Note. From "Managing Improvement by Profiling" by R.E. Blum and J. A. Butler, *Educational Leadership*, 1985, *42*, p. 56.

Step Four -- Goal Setting

The evaluation of the summary narratives by staff members and the school improvement team will lead to the development of a list of areas which need improvement and should be addressed by the school improvement effort. All staff members will have input either in an informal way, through discussion, or in a more formal way by using an evaluation instrument like the one shown in Figures 3 and 4 (Blum &

Butler, 1985). In the formal procedure a rating scale is established to help prioritize the summary narratives. Improvement goals will be established through this process.

Figure 3 -- Example of a Profile Evaluation Worksheet

AREA						
Achievement	Behavior	Attitude	NARRATIVE	Satisfaction Rating	Importance Rating	Acceptable Standard
			Overall, 9 percent of students are absent from one class period each day.			
			The highest number of absentees occurs in the final class period of the day.			
			There is a higher percentage of absentees directly before and after lunch than in morning class periods.			
			Overall, more students are absent in the afternoon than in the morning.			

Note. From "Managing Improvement by Profiling" by R.E. Blum and J.A. Butler, *Educational Leadership*, 1985, *42*, p. 58.

Figure 4 -- Goal Setting Worksheet

AREA					PRIORITY			
Achievement	Behavior	Attitude	Goal and Target	Current Results and Standard	High Priority	Moderate Priority	Low Priority	Short-Term Target
			Improve daily student attendance.	Currently, 91 percent of students attend all courses; the standard is 97 percent.				

Note. From "Managing Improvement by Profiling" by R.E. Blum and J.A. Butler, *Educational Leadership*, 1985, *42*, p. 58.

When setting an improvement goal(s) please consider the answers to these questions:

a. *Can more than one or two goals be reasonably addressed with the available resources?* One goal thoroughly addressed will bring about better improvement than multiple goals inadequately addressed. Consider carefully your resources available for school improvement.

b. *Can the goal(s) be accomplished in the specified time?* Time is an important resource that should be allocated to improvement processes based on the need of the goal. Too often, improvement plans are made to fit one standard school year when much more time is needed. If one school year is all that is available, then choose a goal that can be reasonably accomplished in that time. Choose as another goal one that will be accomplished through on-going improvement efforts.

c. *Are intended improvement outcomes measurable?* Just as student learner outcomes must be measurable, so must school improvement outcomes. Part of the implementation phase of the improvement effort will be evaluating the results. Only measurable results can be compared to pre-prescription data to determine if change has indeed taken place.

   You may need to look at the data over time in order to see results. For example, improvement in norm-referenced test scores may not be apparent unless scores from several years are examined and compared.

Next, the school improvement team selects specific practices, derived from research on schooling, which are likely to bring about the needed improvement. One of the most important factors to consider in selecting improvement practices is how well the staff will accept and ultimately implement that practice. An effective staff development

model advocating a teaching methodology well accepted by influential staff members will probably be more readily accepted than one that is less familiar or somewhat controversial. Again, collaboration in choosing improvement practices enhances the staff's perception that there is indeed school-wide ownership for that improvement practice. Perhaps one of the reasons school-based improvement strategies work more effectively than district-based practices is that collaborative efforts come about more easily within the framework of small groups.

A final word about profiling!

Profiling is a comprehensive study of student achievement, behavior, and attitudes for the purpose of initiating some change strategies. This comprehensive, data-based procedure enables decision-makers to focus on documented facts rather than perceptions in determining what improvement needs to be made. The entire profiling process calls for a commitment on the part of staff for a data-based assessment of student outcomes, an acceptance that change must occur, and a willingness to participate in the improvement process.

Summary

In this handbook the project team of Springfield Public Schools has attempted to raise the important issues regarding instructional assessment. Suggestions have been offered, but the reader is reminded that none is intended to be a "pat" answer.

In concluding, we wish to emphasize once more four major points which have served as the bases for this document. We hope that our readers will consider them as they plan their own assessment systems.

1. The assessment of student learning will be most effective when integrated with curriculum and instruction. If this type of assessment does not support curriculum or instruction, it is misapplied.

2. Assessment may be broadly based and serve many constituents. Its greatest impact on school improvement, however, will occur when managed by and participated in by building-level staff.

3. Effective assessment systems contain varied components. No single element is relied upon to produce information for use by decision-makers.

4. A narrowly defined assessment plan, aimed only at meeting the requirements of state mandates, is likely to have little positive long-term impact on a school or school system.

The members of the team of educators responsible for this guide are more than willing to work with other colleagues interested in enhancing instructional assessment. We invite the questions and comments of those colleagues, and we look forward to the chance to work with them in the future.

References

Alaska Department of Education (1986). *Assessment Handbook: A Practical Guide for Assessing Alaska's Students*. Juneau, Alaska.

Blum, Robert E. and Butler, Joyce A. "Managing Improvement by Profiling." *Educational Leadership*, 42:6, 53-58, March, 1985.

Illinois State Board of Education (1987). *Illinois Assessment Handbook* (draft copy), Springfield, Illinois.

GLOSSARY

(This glossary is adapted, with permission, from the glossary contained in The Illinois Assessment Handbook, published by the Illinois State Board of Education, 1987.)

achievement test - A test that measures the extent to which a student has acquired certain knowledge or mastered certain skills.

alignment - The process of assuring that curriculum, instruction and assessment procedures all match each other, and that communication among educators and administrators at all levels within a district is open and functional.

assessment - Any of a number of methods that may be utilized to gather information about the performance of students. A variety of procedures, including tests, can be used.

comprehensive assessment system - A planned program for monitoring the achievement of an organization's students. The program is likely to specify at least: a schedule that indicates the grade levels of students who will be assessed, the subject areas, and the time (month) of administration; the assessment procedures that will be used; and the plans for processing, interpreting, and using the results.

content validity - The extent to which a test matches the content of a given program.

correlation coefficient - A measure of the degree of relationship between two sets of measures for the same group of individuals. Correlation coefficients range from 0.00, indicating a complete absence of relationship, to +1.00 and -1.00, indicating a perfect positive or negative correspondence.

customized test - A test that is designed specifically for a particular school or district. The test may have been developed by a publisher who

maintains a collection of learning objectives and matching assessment items. The test is developed after the district indicates which objectives to assess.

criterion referenced test (CRT) - A test that is designed to provide information on the specific knowledge or skills of a student. The scores on a criterion referenced test have meaning in relation to what the student knows or is able to do.

educational significance - A judgement that test performance, or the differences in test performance of groups, is meaningful or important in practical terms. This term is often contrasted with statistical significance.

empirical data - Data collected through observation or experience. The test scores of local students, for example, are empirical data.

empirical norm dates - The actual dates on which a publisher tested the students in the norm groups. Publishers recommend that schools administer tests on these dates. Testing at other times may mean that students have received more or less instruction than the norm group.

error of measurement - A statistical estimate of the difference between an observed score and the corresponding "true" score (the score that would be obtained if the assessment were perfectly reliable).

evaluation - What one does with the information gathered in an assessment program.

grade equivalent score (GE) - The grade level for which a given score is the real or estimated average.

item analysis - The process of evaluating individual test items with respect to certain characteristics. Item analysis involves determining such factors as the difficulty level and discriminating power of an item.

All such characteristics are then used to judge the overall quality of the item.

item banks - Collections of assessment items. Generally, these are used for constructing tests that measure selected learning objectives. With sufficient numbers of items, multiple test forms that assess the same objectives can be constructed.

item difficulty level - (See "p value")

normal curve equivalent (NCE) - A measurement scale developed for Title I (Chapter I) evaluation requirements. The scale ranges from 1 to 99, with units equal in size across the score range. The equivalence of units makes it more possible to average scores across groups and to aggregate results across tests.

norm group - The sample of students that was given a test in order to estimate how well the student population in general would perform on the measure. A norm group should be as representative as possible of the variation expected within the general population. Key dimensions to be represented in a norm group include ethnicity, socioeconomic status, size of school system, location of system (urban, rural, or suburban), public vs. non-public schools, and geographic region of the country.

norm referenced test (NRT) - A test that is designed to provide information on how well a student performs in comparison to other students. The scores on a norm referenced test have meaning in terms of their relation to the scores of an external reference group (the norm group).

norms tables - Tables presented in test manuals or available from test publishers that show the relationship of different types of scores to one another (e.g., raw scores to percentiles). Tables are usually provided for each test level and time of testing (norms dates) as well as by grade level of the student tested.

out-of-level testing - Administering a test at a level below or above the one generally recommended for a student based on his or her grade level. Such testing is done to accommodate the ability levels of students who are either much above or much below the average of students their age and thus would not be able to demonstrate the knowledge and skills they possess.

percentile rank - An indication of a student's standing in comparison with all students in the norm group who took the same test. Percentile ranks range from a low of 1 to a high of 99. A percentile rank indicates the percent of students who obtained scores equal to or less than a given score.

pilot testing - Trying out a test or item with a small number of students to see if it works before giving it to a large group of students.

p value - An index which signifies the percent of examinees who answered a test item correctly.

raw score - The number of test items answered correctly by a student. Because different tests have different numbers of items, raw scores cannot be compared from one test to another.

reliability - The extent to which a test can be depended upon to provide consistent, error-free information. Reliability is usually reported as a correlation coefficient, with the closer the coefficient is to +1.00, the higher the reliability. Types of reliability commonly reported for tests include test-retest, alternate forms, split half and Kuder-Richardson KR21.

scale score - A score that expresses the results of a particular test for all forms and levels on a single common scale. Scale scores allow comparisons from grade to grade or level to level of a test.

standard deviation - A measure of the spread of scores.

standard error of the mean - The extent to which sample means can be expected to fluctuate, or vary, due to chance.

standard score - A general term referring to any of several types of "transformed" scores. Scores are expressed in terms of standard scores for reasons of convenience, comparability and ease of interpretation. For example, the raw score of two tests can be expressed in comparable terms by using standard scores.

standardized assessment procedure - A clearly described assessment procedure (for example, a test) with administration directions which were developed so that everyone will administer the procedure in the same way. Student performance will not vary because, for example, administrators give different directions or allow differing lengths of time.

standardized direct writing sample - Short compositions written in response to a writing prompt and scored according to a standardized procedure.

stanine - A standard score scale that ranges from a low of 1 to a high of 9, with a specified percentage of cases falling into each category.

statistical significance - A judgement, based on the application of statistical calculation, that a certain test score or the difference in scores between separate groups are "really" different--that is, not just apparently different because of chance fluctuations. While statistical significance gives the appearance of scientific truth, it must be understood that results of statistical analyses are very dependent on the number of students tested. The smaller the number of scores analyzed, the bigger the difference is required for it to be statistically significant. For this reason, many persons talk about both statistical and educational significance when referring to test scores.

test - ". . . the presentation of a specific set of questions to be answered, a task to be performed, or a problem to be resolved. A test is tangible and

structured, and it can be administered within a relatively limited period of time." (Alaska Department of Education, 1986)

test specifications - Descriptions of the distribution of items for a test. These distributions are frequently used during test construction to specify the number or percent of items that assess various content categories.

testwiseness - The possession of skills that are independent of subject matter knowledge but make it possible for students to achieve higher test scores. Such skills can be taught and will result in small but consistent improvement in test scores.

uniform rating scale - Standards and criteria for assigning value to, for example, student performance. The purpose of a uniform rating scale is to ensure that values will be consistent across raters. A student's rating would be the same regardless of who did the rating.

validity - The characteristic of a test or other assessment procedure that refers to whether the items in an instrument can provide information that is needed for a particular purpose. Most instruments have multiple validities, one for each likely use of the resulting data.

References

Alaska Department of Education (1986). *Assessment Handbook: A Practical Guide for Assessing Alaska's Students.* Juneau, Alaska,